

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2002, 22:13:59 ; Search time 57.04 Seconds

(Without alignments)  
564.899 Million cell updates/sec

Title: US-09-578-063-47  
Sequence: 2247

1 MLETLRQWIVSHRAEWL.....IHLMOEETNLGRCENVL 423

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database:

1: pirl:.\*  
2: pirl:.\*  
3: pirl:.\*  
4: pirl:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1290	57.4	399	2	lysosomal acid lip
2	1290	57.4	399	2	lysosomal acid lip
3	1164.5	51.6	395	1	triacylglycerol 11
4	1139	50.7	397	1	triacylglycerol 11
5	1119	49.8	398	2	triacylglycerol 11
6	752.5	33.5	403	2	hypothetical prote
7	751	33.4	411	2	hypothetical prote
8	730	32.5	405	2	hypothetical prote
9	655.5	29.2	426	2	hypothetical prote
10	574.5	25.6	559	2	egg-specific prote
11	505	22.5	443	2	triglyceride lipas
12	478.5	21.3	1585	2	hypothetical prote
13	475	21.1	467	2	triglyceride lipas
14	445	19.8	344	2	probable lysosomal
15	422.5	18.8	548	2	probable triacylg
16	412.5	18.4	460	2	probable triacylg
17	411.5	18.3	413	2	probable triacylg
18	394	13.1	431	2	protein F15H18.6
19	285	12.7	538	2	probable membrane
20	271.5	12.1	573	2	probable membrane
21	234.5	10.4	509	2	lipase, pregastric
22	149	6.6	62	2	probable esterase/
23	143	6.4	336	2	hypothetical prote
24	136	5.6	987	2	conserved hypotet
25	113	5.0	412	2	triacylglycerol 11
26	112.5	5.0	40	2	hypothetical prote
27	105	4.7	430	2	probable protease
28	103.5	4.6	798	2	Utr protein - Eme
29	101.5	4.5	929	2	

30	101	4.5	436	2	JC5021	platelet-activatin
31	99.5	4.4	286	2	D75217	probable 2-acetyl-
32	99.5	4.4	1280	2	B34087	hypothetical prote
33	99	4.3	460	2	G96764	unknown protein F2
34	97.5	4.3	322	2	B70957	probable epia prot
35	97.5	4.3	466	2	E69391	hypothetical prote
36	96.5	4.3	287	1	E64053	hypothetical prote
37	96.5	4.3	1275	2	I38588	reverse transcript
38	96	4.3	809	2	S67153	probable membrane
39	96	4.3	934	2	B82364	DNA polymerase I V
40	95.5	4.3	1259	4	GNH011	retrovirus-related
41	95	4.2	973	2	T41201	isolauyl-trna syn
42	94.5	4.2	561	2	G83913	hypothetical prote
43	93	4.1	1275	2	B28096	line-1 protein ORF
44	93	4.1	2311	1	TVCRSR	kinase-related pro
45	92.5	4.1	1275	2	S65824	reverse transcript

#### ALIGNMENTS

##### RESULT 1

S41408

lysosomal acid lipase (EC 3.1.1.-) / sterol esterase (EC 3.1.1.13) precursor - human

C:Species: Homo sapiens (man)

C:Date: 13-Jan-1995 #sequence, revision 13-Jan-1995 #text, change 18-Jun-1999

C:Accession: S41408; A39315; S47187

R:Amels, D.; Merkel, M.; Eckerskorn, C.; Greten, H.

Eur. J. Biochem. 219, 905-914, 1994

A:Title: Purification, characterization and molecular cloning of human hepatic lysoso

A:Reference number: S41408; M01D:94155897

A:Accession: S41408

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-399 <AME>

A:Cross-references: EMBL:X76488; NID:9434305; P1DN:CA54026.1; P1D:9434306

J. Biol. Chem. 266, 22479-22484, 1991

A:Title: Cloning and expression of cDNA encoding human lysosomal acid lipase/choleste

A:Reference number: A39315; M01D:92042192

A:Accession: A39315

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-15, 'P', 17-399 <AND>

A:Cross-references: GB:M74775; NID:9187151; P1DN:AA59519.1; P1D:9187152

R:Du, H.; Gregory, G.A.

submitted to the EMBL Data Library, April 1994

A:Description: Structural conservation of putative functional motifs between mouse an

A:Reference number: S47187

A:Accession: S47187

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-22, 'R', 24-399 <DUH>

A:Cross-references: EMBL:Z11690; NID:9506430; P1DN:CA83495.1; P1D:9506431

C:Gene: GDB:LIPA

A:Cross-references: GDB:120153; OMIM:278000

A:Map position: 10q24-10q25

C:Superfamily: triacylglycerol lipase, lingual

C:Keywords: carboxylic ester hydrolase; glycoprotein

Query Match 57.4%; Score 1290; DB 2; Length 399;

Best Local Similarity 61.0%; Pred. No. 6.1e-106;

Matches 242; Conservative 59; Mismatches 94; Indels 2; Gaps 1;

QY	15	MEMMLIVAVMPORNNVSHMPK--AVDPEAFMNSSEIIHQGYCEFEVATEEGYI 72
DB	1	MKRKFGLVVCIVLWMTSHSGSGKLTAVDPETMANVSEIISWGFESEYLVETEFGYI 60
QY	73	LSVNRIPRGLVOPKRTGSRPVLLQHLVGCASWISNLPNNSIGFTIADAGFDVWNGNS 132
DB	61	LCNRRIPRGRKNSDKGRPVVFLQRLGLADSSNWTNLANSSIGFTIADAGFDVWNGNS 120

1 Lysosomal acid lipase - human  
C:Species: Homo sapiens (man)  
C:Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 18-Jun-1999  
C:Accession: G01416  
R:Du, H  
submitted to the EMBL Data Library, April 1994  
A:Reference number: G06919  
A:Accession: G01416  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-399 <DUXX>  
A:Cross-references: EMBL: U08464; NID: g505052; PIDN: AAB60328.1; PID: g505053  
C:Superfamily: triacylglycerol lipase, lingual  
C:Keywords: glycoprotein

RESULT 3  
 LIRTT  
 triacylglycerol lipase (EC 3.1.1.3) precursor, lingual - rat  
 C/Species: Rattus norvegicus (Norway rat)  
 C/Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 11-Jun-1999  
 C/Accession: A23045  
 R/Doercherty, A.J.P.; Bodmer, M.W.; Angal, S.; Verger, R.; Riviere, C.; Lowe, P.A.; L.  
 Nucleic Acids Res. 13, 1891-1903, 1985  
 A/Title: Molecular cloning and nucleotide sequence of rat lingual lipase cDNA.  
 A/Reference number: A23045; MUID:85215587  
 A/Accession: A23045  
 A/Molecule type: mRNA  
 A/Residues: 1-395 <DOC>  
 A/Cross-references: GB:002309; MID:g56595; PIDN:CAA26179.1; PID:g56596  
 A/Experimental source: strain Sprague-Dawley  
 A/Note: The partial sequence of the mature protein from a different, unspecified stru  
 C/Comment: This acid-stable lipase is secreted by the serous (von Ebner's) glands at  
 C/Superfamily: triacylglycerol lipase, lingual  
 C/Keywords: carboxylic ester hydrolase; glycoprotein; lipid digestion; saliva; serous;  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:19-395/Product: triacylglycerol lipase, lingual #status predicted <PPT>  
 F:33,68,98,184,270/Binding site: carboxylate (Asn) (covalent) #status predicted

```

Query Match Similarity      51.8%; Score 1164.5; DB 1; Length 395;
Best Local Similarity      54.9%; Pred. No. 7.1e-95;
Matches 217; Conservative   69; Mismatches 108; Indels    1; Gaps    1

Oy 17 MWLLILVAVMFO-RANVSVMHPTKADVDEAFANNISEIIIOHGOGYPCSEEVATEDGYILSV 75
      ||||| : : : : : ||| ||||| : : ||||| ||||| |
Db 1 MWLLITTSVISIFPGAHGLFCKLGPNGNPANNNISOMITYWGPCOEYVTEDGTYIGLV 60
      ||||| : : ||||| : : ||||| ||||| : : ||||| ||||| |
Oy 76 NRIPGLVOPKKTSGRPVVLLOHLGVGSAMNLSLPNNSLGFLADAGFDVYMGNRCN 135
      ||||| : : ||||| : : ||||| ||||| : : ||||| ||||| |
Db 61 YRIPGKNNSENIGRGPRVVYLQHGILASATXNIAMLPNNSLAFMLADAGDYWLGNRSN 120
      ||||| : : ||||| : : ||||| ||||| : : ||||| ||||| |
Oy 136 AMSRHKLISLDODEFAFYSDEMARFDLPVAINIILKTQGEKIYYVGYSGGTIMGFLA 195
      ||||| : : ||||| : : ||||| ||||| : : ||||| ||||| |
Db 121 TWSRNIVVYSPDSVFEMAFSDEMAKYDLPAININIVOKTGEEKIHVYHSOGTTIGFLA 180
      ||||| : : ||||| : : ||||| ||||| : : ||||| ||||| |
Oy 196 FSTMPELAKIKIMYPALAPIATVKAKSPGTFFLLPDMIKGLFGKKFEILTQTFRLQ 255
      ||||| : : ||||| : : ||||| ||||| : : ||||| ||||| |
Db 161 FSTNTLTAKKIKTEFTALAPVAATVKTKQSPLKISTIFFLEKLMGKKMFLEPHTFDDPL 240
      ||||| : : ||||| : : ||||| ||||| : : ||||| ||||| |
Oy 256 VIYLCGOVILDQICSNIMILLGFENTNMNMSRASVYAHTLAGTSYONILHMSSQAVNS 315
      : : : : : ||||| : : ||||| ||||| : : ||||| ||||| |
Db 241 GTEVCSREVDLDCSNTLFIEFCGFPPKMLNVSREFDVYLGHNPAGISYODFLHQAQLVNSG 300
      : : : : : ||||| : : ||||| ||||| : : ||||| ||||| |
Oy 316 ELRAIDMGSETRKNELECKNQPIPVARKRVADMTVPYTAAMTGGDWLSNPEDYKMLSEVTNL 375
      : : ||||| : : ||||| : : ||||| ||||| : : ||||| ||||| |
Db 301 KFOAQNMGSPSSOMLHYNOKTPPEPDVASMTVPVAVVWNGNDIADPODVALLPRLSNL 360
      : : ||||| : : ||||| : : ||||| ||||| : : ||||| ||||| |
Oy 376 IYKHNPMAHVDFIWGLDAPHBMVNEIHLMOQE 410
      : : ||||| : : ||||| : : ||||| ||||| : : ||||| ||||| |
Db 361 LFKETLAAYNHDLFIWADAPDEVYNEMISMVED 395
      : : ||||| : : ||||| : : ||||| ||||| : : ||||| ||||| |

RESULT 4
JC4017 triacylglycerol lipase (EC 3.1.1.3) cpe precursor - bovine
N/Alternate names: pregastric esterase
C/Species: Bos primigenius taurus (cattle)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C/Accession: JC4017; S64678
R/Timmermans, M.Y.J.; Teuchly, H.; Kupers, L.P.M.
Gene 147, 259-262, 1994
A/Title: The cDNA sequence encoding bovine pregastric esterase.
A/Reference number: JC4017; MUID:95011625
A/Accession: JC4017
A/Molecule type: mRNA
A/Residues: 1-397 <TM>
A/Cross-references: GB:126319; NID:g600756; PIDN:AAAS7037.1; PID:g600757
A/Experimental source: tongue

```

alignment block:

US-09-578-063-47 x HSLAL

Align seg 1/1 to: HSLAL from: 1 to: 2626

2 LeuGluThrLeuSerArgGlnTrpIleValSerHisArgMetGluMetTr 18  
 113 CTGACACACAGCCGCCGACG.....GACAGCTCCAGAAATGAAATGCG 156  
 18 pLeuLeuIleLeuValAlaIleMetPheGlnArgAsnValAsnSerValH 35  
 157 GTCTTGGGGTGTGGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGAG 206  
 35 IsMetProthTrpLys.....AlaValAspProGluAlaPheMetAsnIle 49  
 207 GGCTGGAGGAGAACTGACAGCTGTGTGTCTGTAAACAAACATGAAAGTG 256  
 50 SerGluIleIleGlnHisGlnGlyTyrProCysGluGluIleValAla 66  
 257 AGTGAATATATCTCTACTGTGGGATGCCCTAGTGAGGAATACCTAGTTGA 306  
 66 arhGluAspGlyTyrIleLeuSerValAsnArgIleProArgGlyLeuV 83  
 307 GACAGAAAGATGATATATCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGAG 356  
 83 alGlnProLysLysThrGlySerArgProValIleLeuGlnHisGly 99  
 357 AGAACCATTTGTGACAAAGGTCCCAACCACTGTCTGTCTGTCTGTCTGTCTG 406  
 100 LeuValGlyValAsnSerAsnTrpIleSerAsnLeuProAsnAsnSerLe 116  
 407 TTGCTGGACAAATCTAGTACTGGTGCACAAACCTTCCCAACAGCAGCT 456  
 116 uGlyPheIleLeuAlaAspAlaGlyPheAspValTrpMetGlyAsnSera 133  
 457 GGGCTTCATCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTG 506  
 133 rGlyLysAlaIleTrpSerArgLysHisLysThrLeuSerIleAspGln 149  
 507 GAGGAATACCTGTGTCTGTGAAACATAGACACTGTCTGTCTGTCTGTCTG 556  
 150 GluPheTrpAlaPheSerTyrAspGluMetAlaAspPheAspLeuProAl 166  
 557 GAATTCGGGCTTTCAGTTATGATGAGATGGCAAAATATGACCTACACAGC 606  
 166 aValIleAsnPheIleLeuGlnLysThrGlyGlnGlyLysIleTyrTrp 183  
 607 TTCATTAACCTTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTG 656  
 183 alGlyTyrSerGlnGlyThrThrMetGlyPheIleAlaPheSerThrMet 199  
 657 TGGGTCTCTGTCAAGGACACCATATAGTTTATAGCATTTTTCACACAGATC 706  
 200 ProGluLeuAlaGlnLysIleLysMetTyrPheAlaLeuAlaProIleAl 216  
 707 CCGAGAGTGGCTAAAGGATTAAGTTTTCCTGCTGTCTGTCTGTCTGTCTG 756  
 216 arhValLysHisAlaLysSerProGlyThrLysPheLeuLeuLeuPro 233  
 757 TTCGTCGCGCTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTG 806  
 233 spMetIleIleLysGlyLeuPheGlyLysGlyGluPheLeuTyrGlnThr 249  
 807 ATCATTCATTAAGGACTTATTTGGAGACAAAGAAATTTCTCCCAAGT 856  
 250 ArgPheLeuArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAs 266  
 857 GCCTTTTGAAGTGGGTGGGTACCAAGTTCACATCATCTCATACAGAA 906  
 266 pGlnIleCysSerAsnIleMetLeuLeuGlyGlyPheAsnThrAsnA 283  
 907 GGAGCTCTGTGAAATCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTG 956

283 smkEtsnMetSerArgAlaSerValTyrAlaIleHisThrLeuAlaGly 299  
 957 ATTTAAATATATGTCTAGATGTGATATATACACACATCTCTCTGTCTG 1006  
 300 ThrSerValGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGly 316  
 1007 ACTTCTGTGAAACATGTTACCTGTGAGCAGCAGCTGTAAATTCACAAA 1056  
 316 uLeuArgAlaPheAspTrpIleSerGlyThrLysAsnLeuGluLysCysA 333  
 1057 GTTTCAGACCTTGTAGCTGGGAGAGCAGTGCACAGAAATTAATTCATAC 1106  
 333 snGlnProthTrpProValArgTyrArgValArgAspMetThrValProth 349  
 1107 ACCAGATTATCTCCACATACAAATGTGAAGACATGCTTGTGGCAGCT 1156  
 350 AlaMetTrpThrGlyGlnAspTrpLeuSerAsnProGluAspVal 366  
 1157 GCAGTCTGGAGCGGGGTCTCAGCTGTGCTGTGAGATGTCTACAGCTCA 1206  
 366 smEtsLeuSerGlyValThrAsnLeuIleTyrHisLysAsnIleProG 383  
 1207 TATCTACTACTACATACATACCAACTGTGTGTCTGTGTGTGTGTGTGTG 1256  
 383 IuTrpAlaHisValAspPheIleTrpGlyLeuAspAlaProHisArgMet 399  
 1257 AATGGAGCATCTTGTACTTATTTGGGCTGTGTGTGTGTGTGTGTGTGTG 1306  
 400 TyrAsnGluIleIleHisLeuMetGlnGln 409  
 1307 TATTAATTAATTTATTAATTAATGAGGAA 1336

seq\_name: gb\_pr: H0MLIPCHL

seq\_documentation\_block:

LOCUS H0MLIPCHL 2493 bp mRNA PRI 07-JAN-1995

DEFINITION Human lysosomal acid lipase/cholesterol esterase mRNA, complete cds

ACCESSION M74775.1 GI:187151

VERSION M74775.1

KEYWORDS lysosomal acid lipase/cholesterol esterase.

SOURCE Homo sapiens fibroblast CDNA to mRNA.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Anderson, R.A. and Sando, G.N.

TITLE Cloning and expression of cDNA encoding human lysosomal acid lipase/cholesterol ester hydrolase. Similarities to gastric and lingual lipases

J. Biol. Chem. 266 (33), 22479-22484 (1991)

FEATURES

source

1. 2493 Location/Qualifiers

organism="Homo sapiens"

db\_xref="taxon:9606"

map\_type="Fibroblast"

map="10"

1. 2493

gene="LIPA"

note="G00-120-153"

1. 2493

gene="LIPA"

41. 1240

gene="LIPA"

EC\_number="3.1.1.13"

note="alternate initiator ATG may be at position 47"

codon\_start=1

db\_xref="GDB:G00-120-153"

product="lysosomal acid lipase/cholesterol esterase"

protein\_id="AAA59519.1"

db\_xref="GI:187152"









GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

nucleic search, using sw model

January 30, 2002, 20:27:03 ; Search time 90.11 Seconds  
(without alignments)  
3189,442 Million cell updates/sec

US-09-578-063-46

1 atgttggaaccttgcacag.....gacggtgtgagccgcatgtg 1269

score:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

ed: 351203 segs, 113238999 residues

number of hits satisfying chosen parameters: 702406

num DB seq length: 0  
num DB seq length: 200000000

l-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Issued Patents, NA: \*
- 1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*
  - 2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*
  - 3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*
  - 4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*
  - 5: /cgn2\_6/ptodata/2/ina/PCRNUS.COMB.seq:\*
  - 6: /cgn2\_6/ptodata/2/ina/backfltest1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	444.2	35.0	1137	1	US-08-227-108-2
2	444.2	35.0	1137	2	US-09-073-674-2
3	444.2	35.0	1140	1	US-08-227-108-4
4	444.2	35.0	1140	2	US-09-073-674-4
5	444.2	35.0	1146	1	US-08-227-108-6
6	444.2	35.0	1146	2	US-09-073-674-6
7	444.2	35.0	1528	1	US-08-227-108-1
8	444.2	35.0	1528	2	US-09-073-674-1
9	48.2	3.8	178	1	US-08-751-782-2
10	48.2	3.8	178	2	US-08-925-171-2
11	46.4	3.7	43676	3	US-09-356-952-12
12	36.6	2.9	3271	2	US-08-852-806-1
13	36.6	2.9	3271	3	US-09-163-669-1
14	36.6	2.9	3282	1	US-08-276-852-154
15	36.6	2.9	3282	1	US-08-276-852-154
16	36.6	2.9	3282	1	US-08-276-852-154
17	36.6	2.9	3282	1	US-08-899-575-154
18	36.6	2.9	3282	1	US-08-899-575-154
19	36.6	2.9	3282	1	US-08-899-575-154
20	36.6	2.9	3282	5	PCT-US95-08743-154
21	36.6	2.9	3282	5	PCT-US95-08743-154
22	36.6	2.9	13254	1	US-08-276-852-156
23	36.6	2.9	13254	1	US-08-276-852-156
24	36.6	2.9	13254	1	US-08-899-575-156
25	36.6	2.9	13254	1	US-08-899-575-156
26	36.6	2.9	13254	1	US-08-899-575-156
27	36.6	2.9	13254	1	US-08-899-575-156

28	36.6	2.9	13254	5	PCT-US95-08743-156	Sequence 156, App	
c	29	36.6	2.9	13254	5	PCT-US95-08743-170	Sequence 170, App
c	30	33.6	2.6	8392	1	US-08-080-255-6	Sequence 6, Appl
c	31	33.6	2.6	8392	1	US-08-465-713-6	Sequence 6, Appl
c	32	33.6	2.6	8392	5	PCT-US93-05857-6	Sequence 6, Appl
c	33	32.6	2.6	7218	1	US-08-232-463-14	Sequence 14, Appl
c	34	32	2.5	2197	1	US-08-233-005-3	Sequence 3, Appl
c	35	32	2.5	2197	1	US-08-428-943-3	Sequence 3, Appl
c	36	32	2.5	2197	3	US-09-016-649-3	Sequence 3, Appl
c	37	32	2.5	2197	5	PCT-US95-04858-3	Sequence 3, Appl
c	38	32	2.5	8342	3	US-08-545-860D-63	Sequence 63, Appl
c	39	32	2.5	8342	5	PCT-US94-04496-63	Sequence 63, Appl
c	40	31.6	2.5	8343	4	US-08-965-762-19	Sequence 19, Appl
c	41	31	2.4	3212	3	US-08-673-814-1	Sequence 1, Appl
c	42	30.4	2.4	1529	1	US-08-336-778-1	Sequence 1, Appl
c	43	30.2	2.4	4982	3	US-08-699-103B-1	Sequence 1, Appl
c	44	30.2	2.4	10079	2	US-08-476-866-20	Sequence 20, Appl
c	45	30	2.4	3639	2	US-08-737-524B-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1  
US-08-227-108-2  
Sequence 2, Application US/08227108.  
Patent No. 5807726

GENERAL INFORMATION:  
APPLICANT: Blanchard, Claire  
APPLICANT: Benicourt, Claude  
APPLICANT: Junien, Jean-Louis  
TITLE OF INVENTION: Recombinant Dog Gastric Lipase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/227,108  
FILING DATE: 03-APR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Faunciel, Allan A.  
REGISTRATION NUMBER: 30,256  
REFERENCE/DOCKET NUMBER: 7620-033  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1137 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1137  
US-08-227-108-2

Query Match 35.0%; Score 444.2; DB 1; Length 1137;  
Best Local Similarity 62.8%; Pred. No. 1.5e-134;  
Matches 689; Conservative 0; Mismatches 408; Indels 0; Gaps 0;

```

OY 125 acccagaagcattcatgaatatagtgaaatcatccaacaagaagctccctggaag 184
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 26 ACCCTGAAGTGCACATGAATATAGTCACATGATCCTACTGGGGATACCCAGCTGAGG 85
OY 185 aatatgaagtcgaactgaagaaggatatactctctgttaacaagattccctgagacc 244
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 86 AATATGAAGTTGTGACCGAAGAGGTTATATCCTTGAGTCACACAAATTCCTTATGGA 145
OY 245 tagtgaactgaagaagaagggttccaggccctgtgttactgacgaatgacctagtgtg 304
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 146 GGAATAATTCAGAGATATAGCCCGAGACCTGTTCATTTTTCGACACAGGTTTCTCG 205
OY 305 gaggtgctagcaactggaattcccaactgcccacaatagctggcttattctggcag 364
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 206 CATCAGCCCAACAACTGATCTCCAACTCTCCCAACACAGCCCTGGCTTCACTCTGGCCG 265
OY 365 atgctggttttgacgtgtgtgagtgaggaacagcaggggaaacgcctgtctcgaacaaca 424
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 266 ACGCCGGGTACGACGTGTGGTGGGAAACAGCAGGGGCAACACTGGGCGCAGAGGAAATC 325
OY 425 agaacctctccataagaacgaatgagttctgggcttccagttatgagatgagtagt 484
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 326 TGTACTACTCCGCCCACTCCGTAATCTGGGGCTTTCAGCTTTCAGATGGCTTAAT 385
OY 485 ttgacctctcagtgatgaataacttatttgagaanaacgggccaagaaagatctatc 544
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 386 ATGACTCTCCGCCCACTGACTTCATCTTGAAGAAAGGAGACAGGACGATACACT 445
OY 545 atgtcgtattcaagagggcacaccatgaggtcttattgcatctccacacatgccaagac 604
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 446 ACGTTGGCCATTCAGAGGACACCATGTTTCATCCCTTTTCCACCAATCCCAAGC 505
OY 605 tggctcagaanaatacaaatgtatcttcttgatgaacccaatgccaactgttaagcatgcaa 664
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 506 TGGCAAAACGGATCAAAACCTTCTATGCAATGATGCTCCGCTTCCACCGGAGATGACACG 565
OY 665 aaagcccgaggacaattttgtgtcgcagagataatgagatgaaggatgtgttgga 724
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 566 AAACCTGTATAACAACTCATGCTGCTCCCTTCTCTCAAGCTTATATTTGGAA 625
OY 725 aaaaaaatttctgtatcagaacgaatttctcagaacattgttatttactctgtgccc 784
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 626 ACAAATATTTCTACCCACACACCTTTGATCATTTTCTGCCACCGAGATATGCTCCC 685
OY 785 aggtgattcttgatcagattgtagtaatatcatcaatcttactctgtgtgtgattcaacaca 844
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 686 GCGAGACGCTGATCTCCTCTGACCAAGCCCTGTTATCATTTGTGATTTGACACTA 745
OY 845 acaataigaacatgagcagaatgatatatgctgcccacactcttctgtgaacatctg 904
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 746 TGAACCTTGAACATGATGCTTGGATGTATCTGTACATATATCAGAGAACATCTCG 805
OY 905 tgcanaatatctacactgagcagcagcagtgatctctgtgtaactccggagattgact 964
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 806 TTCAAGAGCTGTCCACAGCTGCTTGAATCTGTGAGGAAAGTTCCAAAGCTTTTGACT 865
OY 965 gggggagtgagacccaanaacttgaanaaataatgcaatcagcacaactccgttaagtgagag 1024
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 866 GGGGAGGCCAGTTCAGAACATGATGACATATCATCAGACGATGCTCCCTCTCTCAACCC 925
OY 1025 tcaagaatatgacgttccctcagaagaatgtagacagagagtgtagacagcttcaaatc 1084
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 926 TGACAGACATGATGTGCCAATGCGAGTGTGAGACGAGTGGCAACGATTTGCGCCGACC 985
OY 1085 cagaagacgtgaaanaatgtgtctctctgtgagtgagacacactcaccataaagaatattc 1144
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 986 CTCACGATGTGACCTTTTGGTTCCAAAGCTGCCCAATCTCATTTACCAACAGAGATTC 1045
OY 1145 ctgaatggtctcaagtgatctatctctgtgtgtgtgtgtgtctctcaacgtatgatacaatg 1204
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1046 CTCCTTAAATCACTTGTGACTTATCTGGGCCATGATGCCCTCAACGCGTTTTCATATG 1105

```

```

OY 1205 aatcatcatcattgattg 1221
DB 1106 AATATGTTTCATGATG 1122

RESULT 2
US-09-073-674-2
; Sequence 2, Application US/09073674
; Patent No. 5998189
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Warner-Lambert Company
; STREET: 2800 Plymouth Road
; CITY: Ann Arbor
; STATE: Michigan
; COUNTRY: U.S.A.
; ZIP: 48105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073, 674
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Crissey, Todd M.
; REGISTRATION NUMBER: 37,807
; REFERENCE/DOCKET NUMBER: 5072-D1-66-TMC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 734 622-7530
; TELEFAX: 734 622-1553
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1137 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1137
; US-09-073-674-2

Query Match 35.0%; Score 444.2; DB 2; Length 1137;
Best Local Similarity 62.8%; Pred. No. 1.5e-134;
Matches 689; Conservative 0; Mismatches 408; Indels 0; Gaps 0;
OY 125 acccagaagcattcatgaatatagtgaaatcatccaacaagaagctccctggaag 184
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 26 ACCCTGAAGTGCACATGAATATAGTCACATGATCCTACTGGGGATACCCAGCTGAGG 85
OY 185 aatatgaagtcgaactgaagaaggatatactctctgttaacaagattccctgagacc 244
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 86 AATATGAAGTTGTGACCGAAGAGGTTATATCCTTGAGTCACACAAATTCCTTATGGA 145
OY 245 tagtgaactgaagaagaagggttccaggccctgtgttactgacgaatgacctagtgtg 304
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 146 GGAATAATTCAGAGATATAGCCCGAGACCTGTTCATTTTTCGACACAGGTTTCTCG 205
OY 305 gaggtgctagcaactggaattcccaactgcccacaatagctggcttattctggcag 364
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 206 CATCAGCCCAACAACTGATCTCCAACTCTCCCAACACAGCCCTGGCTTCACTCTGGCCG 265
OY 425 agaacctctccataagaacgaatgagttctgggcttccagttatgagatgagtagt 484
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

37



```

1007 TTTCGAAGCTCCCAATTCATTACACAGGAAGATTCCTTACAT 1056
386 HisValAspHeIleTrrGlyLeuAspAlaProHisArgMetYTrasnG1 402
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1057 CACTTGGACTTTATCTATGGGCGCATGGATGCCCTTCACGCGTTTCAATGA 1106
402 uileIleHisLeuMet 407
||||:|||||:
1107 AATTGTTTCATGATG 1122

seq.name: /cgn2-6/plodata/2/lna/5B_COMB.seq:US-09-073-674-2

seq documentation block:
Sequence 2, Application US/09073674
Patent No. 5998189
GENERAL INFORMATION:
APPLICANT: Blanchard, Claire
APPLICANT: Benicourt, Claude
APPLICANT: Junien, Jean-Louis
TITLE OF INVENTION: Recombinant Dog Gastric Lipase
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Warner-Lambert Company
STREET: 2800 Plymouth Road
CITY: Ann Arbor
STATE: Michigan
COUNTRY: U.S.A.
ZIP: 48105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073,674
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Crissey, Todd M.
REGISTRATION NUMBER: 37,807
REFERENCE/DOCKET NUMBER: 5072-D1-66-TMC
TELECOMMUNICATION INFORMATION:
TELEPHONE: 734 622-7530
TELEFAX: 734 622-1553
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1137 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1137
US-09-073-674-2

alignment_scores:
Quality: 1141.50 Length: 372
Ratio: 3.624 Gaps: 3
Percent Similarity: 84.677 Percent Identity: 56.452

alignment_block:
US-09-578-063-47 x US-09-073-674-2 ..

Align seg 1/1 to: US-09-073-674-2 from: 1 to: 1137

37 ProHrlySAIaValAspProGluAlaPheMetAsnIleSerGluIle1 53
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
19 CCCACA.....AACCTGAAGTGCATGATATATAGTCAGATGAT 59
53 eGlnHisGlnGlyTrpProCysGlnGluIuTrGluValAlaIarHrgLuAspG 70

```

CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Cisseey, Todd M.  
REGISTRATION NUMBER: 37, 807  
REFERENCE/DOCKET NUMBER: 5072-DI-66-TMC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 734 622-7530  
TELEFAX: 734 622-1553  
TELEX:  
INFORMATION FOR SEQ ID NO.: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 379 amino acids  
TYPE: amino acid

Patent No. 5807726  
GENERAL INFORMATION:  
APPLICANT: Blanchard, Claire  
APPLICANT: Benicourt, Claude  
APPLICANT: Junlan, Jean-Louis  
TITLE OF INVENTION: Recombinant Dog Gastric Lipase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/727,108  
FILING DATE: 03-APR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fannucci, Allan A.  
REGISTRATION NUMBER: 30,256  
REFERENCE/DOCKET NUMBER: 7620-033  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-0900  
TELEFAX: 212 869-8664/9741  
TELEX: 66141 PENNIE

```

; INFORMATION FOR SEQ ID NO: 3:
;
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 379 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
;
US-08-227-108-3

```

Query Match	50.8%	Score 1141.5	DB 1	Length 379
Best Local Similarity	56.5%	Pred. No. 2.4e-112		
Matches 210; Conservative	67;	Mismatches 90;	Indels 5;	Gaps 3

[illegible]

RESULT 6  
US-09-073-674-3  
Sequence 3, Application US/09073674  
Patent No. 5998189  
GENERAL INFORMATION:  
APPLICANT: Blanchard, Claire  
APPLICANT: Benicourt, Claude  
APPLICANT: Junten, Jean-Louis  
TITLE OF INVENTION: Recombinant Dog Gastric Lipase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wainer-Lambert Company  
STREET: 2800 Plymouth Road  
CITY: Ann Arbor  
STATE: Michigan  
COUNTRY: U.S.A.  
ZIP: 48105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/073.674  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Clissey, Todd M.  
REGISTRATION NUMBER: 37,807  
REFERENCE/DOCKET NUMBER: 5072-D1-66-TMC  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 734 622-7550  
TELEFAX: 734 622-1553  
INFORMATION FOR SEQ ID NO: 3  
SEQUENCE CHARACTERISTICS:  
LENGTH: 379 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DS-09-073-674-3

Query Match	50.8%	Score	1141.5	DB 2	Length	379
Best Local Similarity	56.5%	Pred No.	2.4e-112			
Matches 210	Conservative	67	Mismatches	90	Indels	5
					Gaps	3

Oy	37	PTKAVDPAFNNISEIIIOHOGYPCOEYEVAVTEDODIILSVNRIPGLVQPKRSPVLL	96
Dd	7	PT---NEPVYNNISOMIYWCYPAEEIEVATEDJILCIDIRYGRKNSENIGRRPAPL	63
Oy	97	QHGLVGSANMISLNPUNSLGFLLADAGFDVMMGNSRGNAKRHKLSIDODEFAVSX	156
Dd	64	QHGLASATNMISLNPUNSLAFLLADAGDYWLGNCSNGTMWARRNLLYSPDSVEFAFSF	123
Oy	157	DEMAFEDLPVAVINFILOKTGOEKIYYGYISOGTIMGFIASFYMBELAQKIKMYFALAPIA	216
Dd	124	DEMAKYDLPADIDEILKKTGQDKLHYGHSGGTIGFIAFSTNPKLAKRIKTFYALAPVA	183
Oy	217	TYKNAKSPGTEFLLELPMIMIGLEGRKKEPLYQRFELNO-LVYILCGOVILDOIJSNMLL	275
Dd	184	TYKTEYTELLNKMLVPSFLEKFLIRIGNKIF-YPHHFFDQFLATEVCSRETYDILCSNALFT	242
Oy	276	LGGEFTNNMANSRASVYAAHFIAGTSYONILHMSQAVNSGELRAFDESGETKMLEKONP	335
Dd	243	ICGFDTNMLNNSRDLVYLSHNPATSYOVNVLHMSQAKSCKRQAEDMGSPVQNMNHHQS	302
Oy	336	TPVRIYVADMTVPFAAMTGGODMLNSPEDYKMLISEVTNLITYKNNIPEMAHVDFIMGIDA	395
Dd	303	MPPYUNLTDHMYHPIAVVNGNDLLADBDYDILLSKLPNLIYHRKIPPYNHDFIMWADA	362
Oy	396	PHRMVNEIHLML 407	
Dd	363	FOAYVNEIIVSM 374	

RESULT 7  
 US-08-227-108-5  
 Sequence 5, Application US/08227108  
 Patent No. 5807726  
 GENERAL INFORMATION:  
 APPLICANT: Blanchard, Claire  
 APPLICANT: Benicourt, Claude  
 APPLICANT: Junten, Jean-Louis  
 TITLE OF INVENTION: Recombinant Dog Gastric Lipase  
 NUMBER OF SEQUENCES: 21  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Pennine & Edmonds  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: U.S.A.  
 ZIP: 10036  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/227,108  
 FILING DATE: 03-APR-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Faunccl, Allan A.



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

nucleic search, using sw model

January 30, 2002, 20:27:03 : Search time 90.11 Seconds  
(without alignments)  
3189.442 Million cell updates/sec

US-09-578-063-46

1269  
1 atgttggaacctgtcaag.....gacgtgtgagcgctatgt 1269

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

ed : 351203 seqs, 113238999 residues

number of hits satisfying chosen parameters: 702406

num DB seq length: 0  
num DB seq length: 2000000000

Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/prodata/2/1na/5a\_COMB.seq:\*  
2: /cgn2\_6/prodata/2/1na/5b\_COMB.seq:\*  
3: /cgn2\_6/prodata/2/1na/5a\_COMB.seq:\*  
4: /cgn2\_6/prodata/2/1na/5b\_COMB.seq:\*  
5: /cgn2\_6/prodata/2/1na/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/2/1na/bsckfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	444.2	35.0	1137	1	US-08-227-108-2
2	444.2	35.0	1137	2	US-09-073-674-2
3	444.2	35.0	1140	1	US-08-227-108-4
4	444.2	35.0	1140	2	US-09-073-674-4
5	444.2	35.0	1146	1	US-08-227-108-6
6	444.2	35.0	1146	2	US-09-073-674-6
7	444.2	35.0	1528	1	US-08-227-108-1
8	444.2	35.0	1528	2	US-09-073-674-1
9	48.2	3.8	178	1	US-08-751-782-1
10	48.2	3.8	178	2	US-08-751-782-2
11	46.4	3.7	43676	3	US-09-356-952-12
12	36.6	2.9	3271	2	US-08-852-806-1
13	36.6	2.9	3271	3	US-09-163-669-1
14	36.6	2.9	3282	1	US-08-276-852-154
15	36.6	2.9	3282	1	US-08-276-852-154
16	36.6	2.9	3282	1	US-08-276-852-154
17	36.6	2.9	3282	1	US-08-899-575-154
18	36.6	2.9	3282	1	US-08-899-575-154
19	36.6	2.9	3282	1	US-08-899-575-154
20	36.6	2.9	3282	5	PCT-US95-08743-154
21	36.6	2.9	3282	5	PCT-US95-08743-154
22	36.6	2.9	13254	1	US-08-276-852-156
23	36.6	2.9	13254	1	US-08-276-852-156
24	36.6	2.9	13254	1	US-08-899-575-156
25	36.6	2.9	13254	1	US-08-899-575-156
26	36.6	2.9	13254	1	US-08-899-575-156
27	36.6	2.9	13254	1	US-08-899-575-156

28	36.6	2.9	13254	5	PCT-US95-08743-156	Sequence 156, App
29	36.6	2.9	13254	5	PCT-US95-08743-170	Sequence 170, App
30	33.6	2.6	8392	1	US-08-080-255-6	Sequence 6, App1
31	33.6	2.6	8392	3	US-08-465-713-6	Sequence 6, App1
32	33.6	2.6	8392	5	PCT-US93-05857-6	Sequence 6, App1
33	32.6	2.6	7218	1	US-08-232-463-14	Sequence 14, App1
34	32	2.5	2197	1	US-08-428-943-3	Sequence 3, App1
35	32	2.5	2197	3	US-09-016-649-3	Sequence 3, App1
36	32	2.5	2197	5	PCT-US95-04858-3	Sequence 3, App1
37	32	2.5	8342	5	US-08-545-8600-63	Sequence 63, App1
38	32	2.5	8342	5	PCT-US94-04496-63	Sequence 19, App1
39	31.6	2.5	3343	4	US-08-965-762-19	Sequence 1, App1
40	31.6	2.4	3212	3	US-08-673-814-1	Sequence 1, App1
41	31	2.4	1529	3	US-08-336-778-1	Sequence 1, App1
42	30.2	2.4	4982	3	US-08-699-1038-1	Sequence 1, App1
43	30.2	2.4	10079	2	US-08-476-866-20	Sequence 20, App1
44	30.2	2.4	10079	2	US-08-476-866-20	Sequence 20, App1
45	30	2.4	3639	2	US-08-737-524B-26	Sequence 26, App1

## ALIGNMENTS

## RESULT 1

US-08-227-108-2  
Sequence 2, Application US/08227108

Patent No. 5807726

GENERAL INFORMATION:

APPLICANT: Blanchard, Claire

APPLICANT: Benicourt, Claude

APPLICANT: Junien, Jean-Louis

TITLE OF INVENTION: Recombinant Dog Gastric Lipase

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESS: Penne & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.75

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/227,108

FILING DATE: 03-APR-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Farnuel, Allan A.

REGISTRATION NUMBER: 30,256

REFERENCE/DOCKET NUMBER: 7620-033

TELEPHONE: 212 790-9090

TELEFAX: 212 869-8864/9741

TELEX: 66141 PENNE

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1137 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1137

US-08-227-108-2

Query Match 35.0%; Score 444.2; DB 1; Length 1137;  
Best Local Similarity 62.8%; Pred. No. 1.5e-134;  
Matches 689; Conservative 0; Mismatches 408; Indels 0; Gaps 0;

OY	125	acccegaagcatcatgaattagttgaaatcaatcaacaaactaaagctatccctggaag	184
Db	26	ACCCGTAAGGTACCATGATTAATTAAGTACAGATGATCACTCTGTGGGATACCCAGCTGAAG	85
OY	185	aatatgaatgcgaactgaatgaatgggtatataccttcttgtaacagatctccctgaagcc	244
Db	86	AATATGAAGTTGTGACCGGAACACGGTATATATCTTGGGATCGACAGATTCCTTATGGGA	145
OY	245	tagtgcgaaccctaaagaagagttccagggccttgggtgtgtatcgcgcgaatgyccttagttg	304
Db	146	GGAAAAATTCAAGAGAAATATAGCCGGAGACCTGTTGATTTTGTGCAACAGGTTTCTGTG	205
OY	305	gaagtgctagaacatgtgatttccaaactgcgcacaacaaatagccctggagctcatctgcag	364
Db	206	CATCAGCCACAACAGTGGATCTCCAACTGGCCCAACAAACAGCTGGCTTATCTGTGGCG	265
OY	365	atgcgtgttttgcagctgtgtgatggggaacagcaagggaagaagccctggtctcgaaaaaca	424
Db	266	ACCGCGGGTACGAGCTGTGGCTGGGGAAACAGACGGGCAACACTGTGGCCAGAGGAATTC	325
OY	425	agacactctcatagacaagaatgagttctcgggcttcagttatgatatgtagatgagctaggt	484
Db	326	TGTACT.:TGGCCCGGACTCCGTCGAATTTCTGGGCTTTCAGCTTTGACAGATGGCTTAAT	385
OY	485	ttgacactctctgcagtgatgataaacttatttgcagaaaaacgggcagagaagaatgattatc	544
Db	386	ATGACCTTCCCGCACCATTCGACTTCATCTTGAAGAAAGGAGACAGCAACACTACT	445
OY	545	atgtcgagctatccaaagggaacccaatgggcttttatgattatgatttccacatgagccagagc	604
Db	446	ACGTTGGCATTTCCCAAGGCAACCACTTGTTGATTCATGCGCTTTTCCACCAATCCCAAC	505
OY	605	tgtcctcagaaaaatcaaaaatgatttcttcttgaagacc.:tagccactgttgaacatgca	664
Db	506	TGGCGAAAACGGATCAAAACCTTTCATTCATTTAGCTCCGTTGCCACGTAAGTACACCG	565
OY	665	aaagcccccggagacaatttttctgtctgcagatatgatgataaaggatgattgltgca	724
Db	566	AAACCCCTGTTAAACAAACTATGCTCTCTCTTCTGTTCCCTTTCAGACCTATATTTGGAA	625
OY	725	aaaaagaattctgtatcagacccgaatttctcagacaactgtgtatttacccttgagcc	784
Db	626	ACAAAATATTTCTACCACACACACTCTTCTTATCAATTTCTCGCACAGAGTATGCTCC	685
OY	785	aggtgatcttgcatacagatttgatgaataatcatcagttaactctggtgtgattcaacacca	844
Db	686	GCGAGACGGTGAGTCTCTCTCGAGCAACCCCTGTTTATCATTTTGTGATTTGACACTA	745
OY	845	acaatatgaacaatgaagcccgagcaagtgatatactgtccacaactcttctgtgaacaatctg	904
Db	746	TGAACCTTGAAATATGATGCTGTGGATGTGATCTGTACACTAATTCACGAGGAACAATCGG	805
OY	905	tgcacaaatattctcaactctgagccagagcaatgatactctgttgcactccggagcatctgact	964
Db	806	TTTCAGAACGTGCTCACTGTGCTCCAGGCTGTTAAGTCTGGGAAGTTCCAAAGCTTTTGACT	865
OY	965	gggggagtgtagacaacaaaactctggaaaaaatgacatacagccaactcctgttaagtaacagag	1024
Db	866	GGGGAACCCCAAGTTCAACAAATATGATGACATCATATGAGACATGCTCCCTCTCTCAACC	925
OY	1025	tcagagaatatgaacgtctccctacagcaaatgtgagacaggaagtcagagatcgtgcttcaatc	1084
Db	926	TGACAGACATGACATGTCCCAATGGCATGTGGGAAGCGTGGCAACGACTTGCTGTGGCGAAC	985
OY	1085	cagagaagcgtgaaaaatgctcgtctctcgaggttgacaaacctcatctcaataaagaatctc	1144
Db	986	CTCAGCATGTTGACCTTTTCTCTTTCCAAAGCTCCCAATCTCATTTTACCAACAGGAAGATTTC	1045
OY	1145	ctgaatggagctcagcagtgattcatcatccggagtttggatgctctc.:ccgtatgacaatg	1204
Db	1046	CTCCTTACATACACTTGACTTATATCTGGGCCATGATGATGCCCTCTCAACGGGTTTTCAAATG	1105

```
OY 1205 aaatcatccatctgatg 1221
    1111 1 11111
Db 1106 AAATGTTCCATGATG 1122
```

## RESULT 2

US-09-073-674-2  
; Sequence 2, Application US/09073674  
; Patent No. 5008100

GENERAL INFORMATION:

APPLICANT: Blanchard, Claire  
APPLICANT: Benicourt, Claude  
APPLICANT: Tupper, Jean-Louis

AFFILIANT: JAMIE, JEAN-LOUIS  
 TITLE OF INVENTION: Recombinant Dog Gastric Lipase  
 NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Warner-Lambert Company

STREET: 2800 Plymouth Road  
CITY: Ann Arbor

STATE: Michigan  
COUNTRY: U.S.A.

```

;      ZIP:  48105
;      COMPUTER READABLE FORM:
;

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
;

```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,674
; FILING DATE:

```

ATTORNEY/AGENT INFORMATION:  
CLASSIFICATION:  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Crissey, Todd M.  
REGISTRATION NUMBER: 37

REGISTRATION NUMBER: 37,80/  
REFERENCE/DOCKET NUMBER: 5072-D1-66-TMC  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 734 622-7955  
TELEFAX: 734 622-1555

TELEX: 2  
; INFORMATION FOR SEQ ID NO: 2

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1137 base pairs

```

```

;
; TYPE: nucleic acid
; STRANDEDNESS: double

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA

```

```

; FEATURE:
; NAME/KEY: CDS

```

LOCATION: 1..1137  
US-09-073-674-2

Answer Match

Query Match  
Best Local Similarity  
Matches 689: Consensus

maldives 009; conserva  
oy 125 acccaaaacattca

27	accagaaagcattca
26	accctgaagtgcacaa
Db	

QY 185 aatatqaadtccaa

Db 86 AATATGAAGTTGTGA

QY 245 tagtgcacctaaga

Db 146 GGAATAATTCAGAGA

QY 305 gagtgcact

Db 206 CATCAGCCACAACCT

QY 365 atgctggtttgacg

365 atgctggttttcacgtgtgatggtggaacagcagggaacgcctggtctcgaaacaca 424

us-09-578-063-47.rni

About: Results were produced by the GenCore software, version 4.5  
Copyright (c) 1993-2000 CompuGen Ltd.

```
Search information block:
Query: US-09-578-063-47
Query length: 423
Database: Issued_Patents_NA.*
Database sequences: 351203
Database length: 113338999
Search time (sec): 90.070000
```

Sequence	Strd Orig	ZScore	EScore Len	! Documentation
/cgm2_6/torodata/2/ina/5A.COMB.seq:US-08-227-108-2 +			1141.50	2461.53

seq_name	seq	US	08	227	106	2
/cgn2_6/ptodata/2/1na/5b.CONB	seq:US-09-334-9181-4	82.50	140.29	3.65	1455	
/cgn2_6/ptodata/2/1na/5b.CONB	seq:US-08-600-9098-11	82.50	140.26	3.66	1458	
/cgn2_6/ptodata/2/1na/5b.CONB	seq:US-08-663-8381-11	82.50	140.26	3.66	1456	
/cgn2_6/ptodata/2/1na/5b.CONB	seq:US-08-483-101-1	82.50	122.44	35.99	5798	
/cgn2_6/ptodata/2/1na/5a.CONB	seq:US-08-348-891A-1	82.50	109.43	191.04	1589	

Patent No. 5807726

CORRESPONDENCE Address:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.

ATTORNEY/AGENT INFORMATION:

REFERENCE/DOCKET NUMBER: 7620-033

```

; INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

```

STRANDEDNESS: double

LOCATION: 1..1137

Quality: 1141.30  
Ratio: 3.624

[illegible]

53 egl<sup>inh</sup>sgl<sup>ngl</sup>tyr<sup>pr</sup>ocys<sup>gl</sup>



CONVENTION INFORMATION:  
 APPLICANT: Blanchard, Claire  
 APPLICANT: Benicourt, Claude  
 APPLICANT: Junien, Jean-Louis  
 TITLE OF INVENTION: Recombinant Dog Gastric Lipase  
 NUMBER OF SEQUENCES: 21  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Pennie & Edmonds  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: U.S.A.  
 ZIP: 10036  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/227,108  
 FILING DATE: 03-APR-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Fanucci, Allan A.  
 REGISTRATION NUMBER: 30,256  
 REFERENCE/DOCKET NUMBER: 7620-033  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212 790-9090  
 TELEFAX: 212 869-8864/9741  
 TELEX: 66141 PENNIE

2:07:2

RESULT 7  
 US-08-227-108-5  
 Sequence 5, Application US/08227108  
 Patent No. 5807726  
 GENERAL INFORMATION:  
 APPLICANT: Blanchard, Claire  
 APPLICANT: Benicourt, Claude  
 APPLICANT: Junten, Jean-Louis  
 TITLE OF INVENTION: Recombinant Dog Gastric Lipase  
 NUMBER OF SEQUENCES: 21  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Pennie & Edmonds  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: U.S.A.  
 ZIP: 10036  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/227,108  
 FILING DATE: 03-APR-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Faunccl, Allan A.



```
alignment_block:
US-09-578-063-47 x HSLAL
```

Align seg 1/1 to: HSLAL from: 1 to: 2626

2 LeuGIITHTHeuSerArgInIprILleValSerHisArgmetGluMetTr 18  
113 CTGACACACAGCGGCCCGGCAG...GACAGCTCCACAAATGAAATGCG 156  
18 pleuLeuIleleuValAlaTyMetPheGlnArgAsnValAsnSerValH 35  
157 GTTCTGGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGG 206  
35 IsMetProthrllys.....AlaValAspProGlnAlaPheMetAsnIle 49  
207 GGTCTGGAGGAAATCGACAGCTGGAGTCTCGAAACAACATGATATGTG 256  
50 SerGluIleIleGlnHisGlnIleTyTrProCysGlnIuTyrgIuValAl 66  
257 AGCGAAATATATCTTACTACGGGGGATTCCTCTAGTAGAAATCCGATTTGA 306  
66 aThrGluAspGlyTyTrIleLeuSerValAsnArgIleProArgGlyLeuV 83  
307 GACGAGAAGATGGATATATCTCTGCTTACCGAATTCCTCATGGAGGA 356  
83 aGlnProLysTyStrngIysSerArgProValIleLeuGlnHisGly 99  
357 AGAACCATTTCTGCACAAAGTCCCAACAACAGTGTCTCTCTCAACATGGC 406  
100 LeuValGlyIValAserAsnTrpIleSerAsnLeuProAsnAsnSerIle 116  
407 TTGCTGGACAGATTCTGATACAGGGGTGCACAAACCTGGCCACACACAGCT 456  
116 uGlyPheIleLeuAlaAspAlaGlyPheAspValTrpMetLysnSerA 133  
133 rGcIysAsnAlaTrpSerArgLysSHLysTyThrLeuSerIleAspGlnAsp 149  
457 GGGCTTCATCTTCTGCGATGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 506  
507 GAGGAATACCTGGTGTGGGAACATAGACACCTCGAGTTCTTCAGGAT 556  
150 GluPheTrpAlaPheSerTyArgSpGluMetAlaArgPheAspLeuProAl 166  
166 aValIleAsnPheIleLeuGlnLysThrGlyGlnGluLysIleTyTrV 183  
607 TTCATTAACTTCTATCTAATAAACAACGCGCCACAAACACATGATATATG 656  
183 aGlyTySerGlnGlyThrTrpMetGlyPheIleAlaPheSerThmet 199  
657 TGGGTCACTTTCAGGCAACCATATAGGTTTATATGCACTTTTCCACAATC 706  
200 ProGluLeuAlaGlnLysIleLysMetTyrrPheAlaLeuAlaProIleAl 216  
707 CCGAGCTGGCTAAAGATTAATTAATGTTTTTGGCCCTGGGTCCGTGGGC 756  
216 aThrValLysHisAlaLysSerProGlyTyThrLysPheLeuLeuLeuProA 233  
757 TTCCTGGTGGCTTGTACTAGCCCATAGCCCAATTAAGACGATTATTCAG 806  
233 sPheMetIleLysGlyLeuPheGlyLysGlnPheLeuTyGlnTr 249  
807 ATCATCTCATTAAGCACTATTATTGGAGACAAAGAAATTTCTTCCCCACAGT 856  
250 ArgPheLeuArgLnuValIleTyTrLeuCysGlyGlnValIleLeuAs 266  
857 GCGTTTTCGAAGGCGTGGGTACCCACAGCTTTCACACTCATGTCATACGAA 906  
266 pGlnIleCysSerAsnIleMetLeuLeuLeuGlnLysTyrrPheAsnTrpAsnA 283  
907 GAGACTCTGGGAAATCTCTGTTTCTCTCTGCTGGATGATTAATAGAGAA 956

```

293 smetAsmMetSerArgAlaSerValTyrAlaIleHisThrLeuAlaGly 299
||||| :::::::::::::::::::: |||||
997 ATTAAATATGTCTAGAGCTGATGTATATACAAACACATCTCTCGCTGGA 1006
300 ThisSerValGlnAsnIleLeuHisThrPserGlnAlaValAsnSerGly 316
1007 ACTCTGTGGCAAAACATGTATACCTGGAGGACAGCTGTTAAATTCGMAA 1056
316 uleuArgAlaPheAspTyrGlySerGluThrLysAsnLeuGluTyrCys 333
1057 GTTCCAAACCTTTGACTGGGGAACACATGCCAAGATTTATTTTCATTACA 1106
333 snGlnProThrProValArgTyrArgValArgAspMetThrValProThr 349
1107 ACCAGAGTTATCTCCACACATACATGTGAAGACATGCTTGTCGCGACT 1156
350 AlaMetThrThrGlnGlnAspTyrPheSerAspProGluAspValVal 366
1157 GCAGCTGTGGACGGGGGCTACGACTGGCTTGCGAAGTCTACAGACTCA 1206
366 smetLeuLeuSerGluValThrAsnLeuIleTyrHisLysAsnIleProG 383
1207 TATCTTACTGACTCAGATCCACCAACTGCTGATCCATGAGAGCATTCGG 1256
383 IuTPraIleHisValAspPheIleTrrGlyLeuAspIleAspArgMet 399
1257 AATGGACCATCTGCACTTCATTGCGGGCGCTGGATGCCCTTGGAAGCTT 1306
400 TyrAsnGluIleIleHisLysMetGln 409
1307 TATATAATAATTTATTAATCTAATGAGAAA 1336
seq_name: gb_pr:HUMLIPCHL

```

```

seq.documentation block: 2493 bp mRNA PRI 07-JAN-1995
LOCUS HUMLI.PCBL 2493 bp mRNA
DEFINITION Human lysosomal acid lipase/cholesterol estererase mRNA, complete
ACCESSION CDS.
VERSION M74775.1 GI:187151
KEYWORDS lysosomal acid lipase/cholesterol estererase.
SOURCE Homo sapiens fibroblast CDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 2493)
AUTHORS Anderson, R.A. and Sando, G.N.
TITLE Cloning and expression of cDNA encoding human lysosomal acid
lipase/cholesterol ester hydrolase. Similarities to gastric and
lingual lipases
J. Biol. Chem. 266 (13), 22479-22484 (1991)
52042192

Location/Qualifiers
1..2493
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="fibroblast"
/map="10"
<1..2493
/gene="LIPA"
/note="G00-120-153"
1..2493
/gene="LIPA"
41..1240
/gene="LIPA"
/BC_number="3.1.1.13"
/note="alternate initiator ATG may be at position 47"
/codon_start=1
/db_xref="GDB:G00-120-153"
/product="lysosomal acid lipase/cholesterol estererase"
/protein_id="AA059519.1"
/db_xref="GI:187152"

```



```

/translacion="MKRFLGLVYVCLVLPMBHSESGGKLTAVDPETNNVSEIISYV
GFPSEELVETEDGYILCLNRPHPGRKNHSDKPKPVFLQHGILADSSMWVNLANS
SLGFILADAGFVGMGNSRGNTMSRKHTLSQODEMFAISYDEMAKYDDLPASINFL
NKGROVYVYVGHSGCTTIGFIAFSOIPELAKIKMFPAISVAFCSPPAKTIGR
LPDLKLDLFGDKFELPOSFELWLTGHTVHTYIKELCNLCFLICGFERNLNMR
VDYVTHSPAGTISVNNMLHWSQAVKFOKEFQAFMWSGAKNFYFNOSYPPYVANKDL
VPTAVVWSGGHMDLADYDVNILLTQITNLVTFHESIPEMHLEDFIWDADPWRLYNKLI
NMRKRYQ"
41. 103
/gene="LIPA"
/note="alternate signal peptide may begin at position 47;
as predicted from von Heijne, G. (1986) Nucl. Acids Res.
14,4683-4690; G00-120-153"
104. 1237
/gene="LIPA"
/EC_number="3.1.1.13"
/note="G00-120-153"
/product="lysosomal acid lipase/cholesterol esterase"
2466. 2471
/gene="LIPA"
/note="G00-120-153"
2493
polya_site
/gene="LIPA"
/note="G00-120-153"
BASE COUNT 692 a 484 c 516 g 801 t
ORIGIN

alignment_scores:
Quality: 1306.00 Length: 410
Ratio: 3.841 Gaps: 2
Percent Similarity: 82.927 Percent Identity: 60.732

alignment_block:
US-09-578-063-47 x HUMLIPCHL ..
Align seg 1/1 to: HUMLIPCHL from: 1 to: 2493

2 LeuGIuThrLeuSerArgIInTrpIleValSerHisArgMetGluMetTr 18
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8 CTCGAGACAGCGCGCCGCGAG.....GACAGCTCCAGATGAAATAATGCG 51
18 pLeuLeuLeuValAlaIatMetPheGlnArgAsnValAsnSerValH 35
:|||| :|||| :|||| :|||| :|||| :|||| :|||| :|||| :||||
52 GTTCTGGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 101
35 IsMetProThrLys.....AlaValAspProGluAlaPheMetAsnIle 49
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
102 GGTCTGGAGGAACTGACAGCTGTGGATCTGAAACAAACATGATGAG 151
50 SerGIuIleIleGlnHisGlnIleTyProCysGluGluGluValAl 66
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
152 AGTGAATATCTCTACTGCGGATGCCCTAGTGAAGAAATACCTACTTGA 201
66 aThrGluAspGIuTyTrIleLeuSerValAsnArgIleProArgGluLeuV 83
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
202 GACAAAGATGATATATTTCTGCTTACCGAATTCCTCATGGAGAGA 251
83 aGlnProLysLysThrGlySerArgProValValLeuLeuGlnHisGly 99
:|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
252 AGAACCATCTTCACAAAGGTCCCAACAGTGTCTCTCCACAACTGGC 301
100 LeuValGlyGlyAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerIle 116
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
302 TTGCTGGCGAGATCTAGTAAGTGGGTCACAAACCTTGCCAAAGCGCT 351
116 uGlyPheIleLeuAlaAspAlaGlyPheAspValTrpMetGlyAsnSerA 133
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
352 GGGCTTCATCTTGGCTGATGCTGCTTTTACGTGGATGGGCAACAGCA 401
133 rGgLYAsnAlaTrpSerArgLysHisLysThrLeuSerIleAspGlnAsp 149
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
402 GAGGAATAACCTGCTCGGAAACATAAGACACTCTCACTTCTCGAGAT 451

```

```

150 GluPheTrpAlaPheSerTyArgSpGluMetAlaArgPheAspLeuProAl 166
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
452 GAATCTGGGCTTTCAGTATGATGATGATGATGATGATGATGATGATGAT 501
166 aValIleAsnPheIleLeuGlnLysThrGlyGlnGluLysIleTyTyV 183
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
502 TTCCATTAATCTTCATTCGAAATAAACATGGCCAGAACAGATTAATAG 551
183 aGlyTySerGlnGlyThrThrMetGlyPheIleAlaPheSerThrMet 199
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
552 TGGGTCATTCCTAAGGCACACTATGATGATGATGATGATGATGATGATG 601
200 ProGluLeuAlaGlnLysIleLysMetTyPheAlaLeuAlaProIleAl 216
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
602 CCGAGCTGCTTAAAGATTAATGTTTTCCTCCCTGGCCGTGGGC 651
216 aThrValLysHisAlaLysSerProGlyThrLysPheLeuLeuProA 233
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
652 TTCGTCGCTTCTGTAAGCTTATTTGGAGCAAAATTTCTCCAGACT 701
233 sPheMetIleLysGlyLeuPheGlyLysGluPheLeuTyGlnThr 249
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
702 ATCATCTCATTAAGCACTTATTTGGAGCAAAATTTCTCCAGACT 751
250 ArgPheLeuArgGlnLeuValIleTyLeuCysGlyGlnValIleLeuAs 266
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
752 GCGTTTGAAGTGGCTGGGTACCCAGCTTTCACATCATGATACTGAA 801
266 pGlnIleCysSerAsnIleMetLeuLeuGlnGlyPheAsnThrAsn 283
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
802 GAGCTCTGTGGAATCTGTTTCTGTTCTGTTGATTAATGAGAGA 851
283 sMetAsnMetSerArgAlaSerValTyAlaIleHisThrLeuAlaGly 299
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
852 ATTTAATATGCTAGAGTGAATGATATATACAAACATTCCTGCTGGA 901
300 ThrSerValGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGly 316
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
902 ACTTCTGTCAAAACATGTTACCTGAGCCAGCGCTTAAATTCCAAAA 951
316 uLeuArgAlaPheAspTrpGlySerGluThrLysAsnLeuGluLysCysA 333
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
952 GTTCAAGCTTGTGACTGGGGAAGCAGTGCACAAATTTTTCATTACA 1001
333 sGlnProThrProValArgTyArgValArgAspMetThrValProThr 349
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1002 ACCAGATTAATCTCCACATCAATGTAGAGACATGCTGTGGCGACT 1051
350 AlaMetTrpThrGlyGlyGlnAspTrpLeuSerAsnProGluAspVal 366
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1052 GCAGTGTGAGGGGGGTGACAGCTGGCTTGCAAGTGTACGACGTCAA 1101
366 sMetLeuLeuSerGluValThrAsnLeuIleTyHisLysAsnIleProG 383
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1102 TATCTTACGATCAGATCACCACACTGGTGTTCATGAGAGATCCGG 1151
383 IuTrpAlaHisValAspPheIleTrpGlyLeuAspAlaProHisArgMet 399
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1152 AATGGAGCATCTTGACTTCATTTGGGGCTGGAATGCCCTTGAGAGCTT 1201
400 TyrAsnGluIleIleHisLeuMetGlnGln 409
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1202 TATATAATAATTAATTAATTAATGAAGGAAA 1231

seq_name: gb_pr:BC012287
seq_documentation_block:
LOCUS BC012287 mRNA 2586 bp 08-AUG-2001
DEFINITION Homo sapiens, Similar to lipase A, lysosomal acid, cholesterol
esterase (Wolman disease), clone MGC:5229 IMAGE:2900168, mRNA,
complete cds.
ACCESSION BC012287

```

```

VERSION      BC012287.1  GI:15126726
KEYWORDS
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 2586)
AUTHORS      Strausberg, R.
TITLE        Direct Submission
JOURNAL      Submitted (06-AUG-2001) National Institutes of Health, Mammalian
              Gene Collection (MGC), Cancer Genomics Office, National Cancer
              Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
              USA
REMARK       NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT      Contact: MGC help desk
              Email: cgaps-remail.nih.gov
              Tissue Procurement: ATCC
              CDNA Library Preparation: Life Technologies, Inc.
              CDNA Sequencing by: The I.M.A.G.E. Consortium (ILNI)
              DNA Sequencing by: Baylor College of Medicine Human Genome
              Sequencing Center
              Center code: BCM-HGSC
              Web site: http://www.hgsc.bcm.tmc.edu/cdna/
              Contact: villalobebcm.tmc.edu
              Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
              A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
              Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNI at: http://image.llnl.gov
Series: IRAX Plate: 3 Row: c Column: 3
This clone was selected for 'full length sequencing because it
passed the following selection criteria: matched mRNA gi: 434305.

FEATURES
Source
1. 2586
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /clone="MGC:5129 IMAGE:2900168"
   /tissue_type="Placenta, choriocarcinoma"
   /clone_1id="NH_MGC_10"
   /lab_host="DH10B"
   /note="Vector: pCMV-SPORT6"
   /codon_start=1
   /product="Similar to lipase A, lysosomal acid, cholesterol
esterase (Wolman disease)"
   /protein_id="AAH12287.1"
   /db_xref="GI:15126727"
   /translation="MKMRFLGLVCLVLMPLHSESGSGKTLALDPETNMVSEIISYK
GPFSEELVETEDGYILCLNRIPIGRNRHSDKGPVFLQHGILADSSWVYMLANS
SLFLFIADAGFVWGMGRNWTMSRKRTSLSDDEFMFASYDEMAKVDLPAINFIL
NRKGOEOLVVYVGHSGGTIGEIFASQIPELAKRIKMFALGPVAVAFSPAKLGR
LPHLRIELFDGKEFLPQSAFLKMFKMGVCTHYILKICGNCITLCGFMRNIMSR
VDVYTHSPAGTSYQNNLHNSQAVREKRFAPFDGSSAKNFRHNSQTPPTYNKDL
VPIAVMSGHMDLADVDIVNLILQITNLVHFHSIPFWHLDFITGLDAPVRYNKKIIL
NLNRKY0"

BASE COUNT   736 a      511 c      531 g      808 t
ORIGIN
alignment_scores:
  Quality: 1303.00      Length: 410
  Ratio: 3.832          Gaps: 2
  Percent Similarity: 82.927      Percent Identity: 60.488

alignment_block:
US-09-578-063-47 x BC012287
..
Align seg 1/1 to: BC012287 from: 1 to: 2586

2 LengUThleuSerArgGlnTrpIleValISerHisArgMetCJumetr 18
|||||||.....|
68 CTCGAGACACGCGCGCGGAG.....GACACCTCCAGATGATAAATGCG 111

```

[illegible]

[illegible]

JOURNAL Submitted (08-APR-1994) Hong Du, Division of Human Genetics, Children's Hospital Medical Center, 3333 Bernet Street, Cincinnati, OH 45229-3039, USA

FEATURES

source Location/Qualifiers

1. 2481

5'UTR

sig\_peptide

CDS

34..96

34..123

/codon\_start=1

/product="lysosomal acid lipase"

/protein\_id="AAB60328.1"

/db\_xref="GI:505053"

/translation="MKRRLFLGVCLVLMTHLSRSGSKLTAVDPETNMVNSLISW  
GPEPEEYLVETEDGYILCLNTHPRGRNHSQKPKPVFLOHGLADSSNVVNLANS  
SLGILADAFEDVMGNSRGNTWPKRKTLSODEFWAFSDMAKYDLPASINFILIG  
NKGQOEYVVYVGHSGQTIGLAFSOLPELAKRIKMPFALGPVSVAFCSTPMKLGIR  
LPOLHILKIDLFQDKRFLPQSAFLKNGLGHVCHYIKLCLGNCILCGFENRNLMSR  
VDYTHSPAGTSVQNMHSQAVKRFQAFDMGSSAKNFIHNSQPTPYNKKDM  
VPLVMSGSGHMDLVADVIVNLLTQTNLVFHESIPEMHLDFIWDGAPRLYKII  
NLMKRYO"

79

/replace="C"

97..1230

/product="lysosomal acid lipase"

100

/replace="G"

1234..2480

2459..2464

2480

3'UTR

polya\_signal

polya\_site

BASE COUNT 692 a 477 c 513 g 799 t

ORIGIN

alignment\_scores:

Quality: 1298.00 Length: 400

Ratio: 3.875 Gaps: 1

Percent Similarity: 83.750 Percent Identity: 61.000

alignment\_block:

US-09-578-063-47 x HSU08464 ..

Align seg 1/1 to: HSU08464 from: 1 to: 2481

12 SerHISArgMetGluMetTrpLeuLeuLeuValAlaIleTyrMetPheG1 28

|||||.....|

25 AACCTCCAGATGAAATAATGCGGTCTTGCGGTGGTGCTGTTCGTTCTTCT 74

28 nArgAsnValAsnSerValHisMetProThrIys.....AlaValAsp 43

|||||.....|

75 CMGGAACCCGCAATTCGAGGGGCTAGAGGAAACGACAGCGTGCGATC 124

43 roGluAlaPheMetAsnIleSerGluIleIleGlnHisGlnIleTyrPro 59

|||||.....|

125 CTGAACAACAACATGATGATGTAATATATCTTACTACGGGGGATCCCT 174

60 CysGluGluTyrGluValAlaIleThrGluAspGluTyrIleLeuSerValAs 76

|||||.....|

175 AGTGAAGATACCTAGTTGAGACAGAGATGATATATTCGTGCTTAA 224

76 nArgIleProArgGlyLeuValGlnProIlysthrGlySerArgProV 93

|||||.....|

225 CCGAATTCCTCATGGGAGAGAACACATCTGCACAAAGTCCCAAAACAG 274

93 aIValleuLeuGlnHisGlyLeuValGlyAlaSerAsnTrpIleSer 109

|||||.....|

275 TTGCTTCTCTGCACATGCTGCTGGCGGCAATTCAGTAACTGAGGCGACA 324

```

seq_name: gb_ro:S81497
seq_documentation_block: 3144 bp mRNA ROD 02-AUG-2000
LOCUS S81497 lysosomal acid lipase-intracellular hydrolase [rats, Wolman, liver,
DEFINITION mRNA, 3144 nt].
ACCESSION S81497
VERSION S81497.1 GI:1336725
KEYWORDS
SOURCE Rattus sp. liver Wolman.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 3144)
AUTHORS Nakagawa,H., Matsudaira,S., Kuriyama,M., Yoshidome,H., Fujiyama,J.,
TITLE Yoshida,H. and Osame,M.
JOURNAL Cloning of rat lysosomal acid lipase cDNA and identification of the
MEDLINE mutation in the rat model of Wolman's disease
REMARK J Lipid Res. 36 (10), 2212-2218 (1995)
66129534
GenBank staff at the National Library of Medicine created this
entry [NCBI gidsq 176569] from the original journal article.
This sequence comes from Fig. 2.
FEATURES
location/Qualifiers
source
1..3144
/organism="Rattus sp."
/db_xref="taxon:10118"
1..3144
/gene="lysosomal acid lipase, LAL"
233..1426
/gene="lysosomal acid lipase, LAL"
/feature="lysosomal acid lipase, LAL"
/feature="intracellular hydrolase" This sequence comes from
Fig. 2; conceptual translation presented here differs from
translation in publication; LAL"
/codon_start=1
/product="lysosomal acid lipase"
/protein_id="AAB36043.2"
/db_xref="GI:9653291"
/translation="MOLLGRVLCFVVGILSGPRTISAVDPANNVTEIIMHGGT
/translation="MOLLGRVLCFVVGILSGPRTISAVDPANNVTEIIMHGGT
PEHVSOTGDEYILGVHRIPIGRKNOFPGKPPVYLMWRHGFADSSMWTINDNSI
GFILADAGFVMMGNSRGNTWRSRKHTLSODEYNAFSEDEMAKYDLPASINYLNNK
TGOGLANFVNGSOGCTGFIATFASOMPELAKVKEFFALPVLSTNFAFGPVRKGRLE
DILEDIEFGOKRFLPOSAMVWSTFHCITVIMKEICANIFELICGFENKLNNSRP
VYTHHCAGTISVQNMVHWOTVYVKKHKLQADPMGSDKNYFRINOSYPLRSIKRMQPD
TALMSGGKWLADTSDINILITTEPTLVYHKNIPEWHLDEFIWLDPAPMYLINEVPS
MKKYO"
BASE COUNT 852 a 690 c 754 g 848 t
ORIGIN
alignment_scores:
Quality: 1220.50 Length: 378
Ratio: 3.838 Gaps: 3
Percent Similarity: 84.127 Percent Identity: 58.201
alignment_block:
US-09-578-063-47 x S81497 ..
Align seg 1/1 to: S81497 from: 1 to: 3144
37 Prothr.....LysAlaValAspProGluAlaPheMetasrIleSe 50
||||| :|||
290 CCAACGGGACCATATCATCTGACCTGTGACCTGTGAAGCAAAATGATGATGAC 339
||||| :|||
50 xGluIleIleGlnHISgInGlyTyRProCysgInGluIuTyRgIuValAlaT 67
||||| :|||
340 TGAAGATTAATCATGACACTGGGATATFCA.....GAGCACTAGTACAGA 383
||||| :|||
67 hGcIuAspGlyTyRtIleLeuSerValAsnArgIleProArgGlyLeuVal 83
||||| :|||
384 CAGGAGATGGCTACATCTTGCGGTTCACCGGAATCCCGATGGCGGGAAG 433
||||| :|||

```

[illegible][illegible]

alignment\_block:  
US-09-578-063-47 x MMLYACLY

Align seg 1/1 to: MMLYACLY from: 1 to: 2358

[illegible][illegible]

```

98 AACACATGCTCTTTGGAAATGACACCAACA.....AACCC 138
43 rogluAlaPheMetAsnIleSerGluIleIleGlnHisGlnGlyPro 59
139 CTGAAGTGAATATGAAATTTAGTCAGATGATTTCTTACTGGGAGATCCCA 188
60 CysGluGluTyrGluValAlaThrGlnuSpGlyTyrIleLeuSerValAs 76
189 AGTGAATAATATGAAATTTAGTCAGATGATTTCTTACTGGGAGATCCCA 238
76 nargIleProArgGlyLeuValGlnProLysThrGlySerArgProV 93
239 CAGAAATTCCTATGAGGAGAAATAATTCAGAGACAGGAGGAGACCCG 288
93 alValLeuLeuGlnHisGlyLeuValGlyGlyAlaSerAsnTrpIleSer 109
289 TCGTATTTTGCAGCATGTTGCTTGCATCAGCCTCAACTGATCTCC 338
110 AsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAspAlaGlyPheAs 126
339 AACCTGCCCAACACAGCTGCTGCTTTATTTCTGCGAGATGCTGTTATGG 388
126 pValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerArgLysHisLysT 143
389 TGTGTGCTGGGGAACAGCAGAGAAATACCTGCTCCAGAGAAACTTAT 438
143 hrLeuSerIleAspGlnAspGluPheTrpAlaPheSerTyrAspGluMet 159
439 ACTATTCACAGACCTCCCTGTAATTTCTGGCTTTTACCTTGATGATAATG 488
160 AlaArgPheAspLeuProAlaValIleAsnPheIleLeuGlnLysThrGln 176
489 GCTAAATATGACCTTCAGCCACCAATTCATTCATTTGAAGAAACGTGG 538
176 yGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyThrTrpMetGlyP 193
539 ACAGAAAGAGCTGCATATGTTGTCATCCAGGAGCAGCAGCATGGT 588
193 heIleAlaPheSerThrMetProGluLeuAlaGlnLysIleLysMetTyr 209
589 TTATGTGCTTTTTCACAAATTCACCAAGCTGGCTGAAAGATCAAAACCTT 638
210 PheAlaLeuAlaProIleAlaThrValLysHisAlaLysSerProGlyTh 226
639 TATGATTAAGTCCAGTCGACCACTGAGATACAAAAGCCTTGTA 688
226 rLysPheLeuLeuLeuProAspMetMetIleLysGlyLeuPheGlyLysL 243
689 CAACCTTAAGTATTATCTCTCAACACATGTCAGATTTATTTGGTGACA 738
243 ysgLupheLeuTyrGlnThrArgPheLeuArgGlnLeuValIleTyrLeu 259
739 AAATATTCTACCCACACAAATTTCTTGATCAATTTCTTGCCACATCAAGT 788
260 CysGlyGlnValIleLeuAspGlnIleCysSerAsnIleMetLeuLeu 276
789 TGTTCCTCGTGAGACATGATGATTTGTCAGCAATGCTTATTTATAT 838
276 uGlyLysPheAsnThrAsnAsnMetLysMetSerArgAlaSerValTyrA 293
839 TTGTGGCTTTGACAGCGCAACTGGAACATGATGCTGGATGTGTACG 888
293 laAlaHisThrLeuAlaGlyThrSerValGlnAsnIleLeuHisTrpSer 309
889 TGTCCATCAATATCCGCGAGAACTTCAGTTCAAAACATCTGCACATGGACC 938
310 GlnAlaValAsnSerGlyLysLeuArgAlaPheAspTrpGlySerGluTh 326
939 CAGGCTGTTAAATCTGGAATTTTCACCTTTTATTTGGGAGAGCCAGC 988
326 rLysAsnLeuGluLysCysAsnGlnProThrProValArgTyrArgValA 343

```

```

989 TCAGATGTAGTGCACCTTCAATCAGCCACACCTCCCTACTACATGTGA 1038
343 rgaSPmetThrValProThrAlaMetTrpThrGlnGlyGlnAspTrpLeu 359
1039 CCGCATGAATGTGCCAATGTGTGAGGTGGAGGCGCAATGACTGGTGG 1088
360 SerAsnProGluAspValLysMetLeuLeuSerGluValAlaTrpAsnLeu 376
1089 GGTGACCCCAAGAGATGTGACCTTTGCTTCCAAACCTTCAATCTCAT 1138
376 eTyrHisLysAsnIleProGluTrpAlaHisValAspPheIleTrpGlyL 393
1139 TTACCAACAGAGATTCCTCCATATATCACTGATTTATCTGGGCGAA 1188
393 euAspAlaProHisArgMetTyrAsnGluIleIleHisLeuMetGlnGln 409
1189 TGAATGCTCTTCAGAGATTTTACAAATGAAATTTATTTCTATGATGCANA 1238
410 GluGlu 411
1239 GATAAA 1244

```

```

seq_name: gb_pat:A26690
seq_documentation_block: 1378 bp DNA PAT 05-APR-1995
LOCUS A26690
DEFINITION Precursor of rabbit gastric lipase (comp.).
ACCESSION A26690
VERSION A26690.1 GI:905030
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequence.
REFERENCE
1 (bases 1 to 1378)
AUTHORS
Benicourt,C., Blanchard,C. and Junien,J.L.
TITLE
Recombinant gastric lipase from rabbit and pharmaceutical
compositions
JOURNAL
Patent: EP 0542629-A, 10 19-MAY-1993;
INSTITUT DE RECHERCHE JOUVEINAL
location/Qualifiers
1. 1378
source
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT 418 a 276 c 287 g 397 t
ORIGIN

```

```

alignment_scores:
Quality: 1167.00 Length: 402
Ratio: 3.505 Gaps: 3
Percent Similarity: 82.836 Percent Identity: 54.478

```

```

alignment_block:
US-09-578-063-47 x A26690/rev ..

```

```

Align seg 1/1 to reverse of: A26690 from: 1 to: 1378

```

```

16 GluMetTrpLeuLeuIleLeuValAlaTyrMetPheGlnArgAsnValAs 32
1328 AAAATGTGGTGCCTTTTCATGTGTGACAGCTTGTGATCTGCA...CTTGG 1282
32 nserValHis.....MetProThrLysAlaValAsp 43
1281 AACCTACATGCTCTTTTGGAAATGACACCAACA.....AACCC 1241
43 rogluAlaPheMetAsnIleSerGluIleIleGlnHisGlnGlyPro 59
1240 CTGAAGTGAATATGAAATTTAGTCAGATGATTTCTTACTGGGAGATCCCA 1191
60 CysGluGluTyrGluValAlaThrGlnuSpGlyTyrIleLeuSerValAs 76
1190 AGTGAATAATATGAAATTTAGTCAGATGATTTCTTACTGGGAGATCCCA 1141

```



```

76  natgileproargglyleuvalglnprolyserthglyserargprov 93
   ||||| |||
1140 CAGAAATTCCTATGCGAAGAAATTCAGGAAACAGAGCCAGAGACCG 1091
   ||||| |||||
93  alyalleuleuvalglnhslglyleuvalglnlyalaseratpilleser 109
   ||||| |||||
1090 TCGATATTTTTCAGACATGTTGCTTCATCAGCTCAAACTGATCTCC 1041
   ||||| |||||
110  Asnleuproasnserleuglypheiileuvalaspralaglypheas 126
   ||||| |||||
1040 AACCTGCCACACAGACCTGGCTTATCTGCGAAGTGGTGTATGG 991
   ||||| |||||
126  pvaltrpmetglyasnserargglyasnalaatrpseryshlyst 143
   ||||| |||||
990 TGTGTGCTGGGAGACAGACAGAAATFACCTGCTCCAGAAAGAACTTAT 941
   ||||| |||||
143  hrlseuSerileaspGlnaspGlnpheiTrpAlapheSerTyaspGlnMet 159
   ||||| |||||
940 ACTATTCACAGACTCCGTTGAATCTGGGCTTTTACCTTGATGAATG 891
   ||||| |||||
160  AlalargpheaspLeuproAlaValIleasnPheIleuGlnlystngl 176
   ||||| |||||
890 GCTAAATATGACCTCCAGCCACAAATGACTTCATGTAAAGAACTGG 841
   ||||| |||||
176  yGlnGlnlysllytyrtyrvalglytyrserGlnGlythThrmetylP 193
   ||||| |||||
840 ACAGAGAAACCTGACATATGTGTGTCATCCAGGACACCCCTTGTGT 791
   ||||| |||||
193  heilealapheserThrmetrproGlnleuvalaglnlysllyslly 209
   ||||| |||||
790 TTTATGCTTTTCTACAAATCCCAAGCTGGCTGAAGATCAAAACCTT 741
   ||||| |||||
210  PheAlaLeuAlaProIleAlaThValLysHisAlaLysSerProGlyTh 226
   ||||| |||||
740 TATGCAATACCTCCAGTCCGACACTGTAAGTAAACCTTGTAA 691
   ||||| |||||
226  rlyspPheleuleuProaspmetMetIlelysllyleupheGlysl 243
   ||||| |||||
690 CAAACTAGGTTTATTCCTCCAAACATGTCACAAATTAATTTGATGCA 641
   ||||| |||||
243  yslguleuleuTyGlnThrArgPheleuArgGlnleuValIleTyLeu 259
   ||||| |||||
640 AAATATCTCAACCCACAAATTTCTGATCAATCTTCCACACACAGTG 591
   ||||| |||||
260  CysGlyValIleleuaspGlnIleCysSerAsnIleMetLeuLeu 276
   ||||| |||||
590 TGTCCCGTAGACACACTGAATGTAATTCGCAATGCTTATTTATCAT 541
   ||||| |||||
276  uGlyGlyPheAsnThrAsnMetAsnMetSerArgAlaSerValTyG 293
   ||||| |||||
540 TTGTGGCTTTGACAGCGCAACTGAAACATGAGTGGCTGGATGTGACG 491
   ||||| |||||
293  laaAlaThrLeuAlaGlyThrSerValGlnAsnIleLeuHisTrpSer 309
   ||||| |||||
490 TGTCAATATCCGCGACAGAACTGATCAAAACATGCTGCACTGAGAC 441
   ||||| |||||
310  GlnAlaValAsnSerGlyGlnleuAlaAlaPheAspTrpGlySerGln 326
   ||||| |||||
440 CAGGCTGTGTAATCTGAAATTTTCAAGCTTTTAATTCGGGAACTCCAG 391
   ||||| |||||
326  rlyAsnleuGlnlysllyCysAsnGlnProThrProValArgTyArgVal 343
   ||||| |||||
390 TCAAAATGTGTCATCTCAATGACCCACACTCCCTACTACAAATGTGA 341
   ||||| |||||
343  rGaspMetThrValProThrAlaMetTrpThrGlyGlnAspTrpLeu 359
   ||||| |||||
340 CCGCAATGATGTGCAATGCAATGTGAGTGTGCAATGACTGTG 291
   ||||| |||||
360  SerAsnProGlnAspValLysMetLeuLeuSerGlnValThrAsnLeu 376
   ||||| |||||
290 GCTGACCCCAAGATGTGACCTTGTGCTCAAACTTCTTAATCTCAT 241
   ||||| |||||
376  eTyHisIleLysAsnIleProGlnIleThrAlaHisValAspPheIleTrpGly 393

```

```

||||| |||||
240  TTACCACAGAGATCTTCATATATACACTGGATTTATCTGGGCAA 191
   ||||| |||||
393  euAspAlaProHisArgMetTyTrasnGlnIleIleHisLeuMetGln 409
   ||||| |||||
190  TGAATGCTCCCAAGATTTTCAATGAAATTAATTTCTATGATGCAAAA 141
   ||||| |||||
410  GluGlu 411
   ||||| |||||
140  GATAAA 135
   ||||| |||||
seq_name: gb_pat:A01157
seq_documentation block:
LOCUS      A01157          1336 bp      mRNA      PAT      29-JAN-1993
DEFINITION R.norvegicus mRNA for prelingual lipase protein.
ACCESSION  A01157
VERSION    A01157.1 GI:14748
KEYWORDS   lipase.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Ratus.
REFERENCE  1 (bases 1 to 1336)
AUTHORS   Williamson,R.
JOURNAL   Patent: WO 8500381-A 3 31-JAN-1985;
            Celltech Ltd
FEATURES
    source
        1..1336
            location/Qualifiers
                /organism="Rattus norvegicus"
                /db_xref="taxon:10116"
                /codon_start=1
                /product="prelingual lipase"
                /protein_id="CAA00136.1"
                /db_xref="GI:14749"
                /db_xref="SWISS-PROT:P04634"
                /translation="MMLLITVISTFGGAGLFGKLGPNPEANKNISQMTTCYF
                COEEVYTEDGYILGVRIPIHKNNSENIGKRPVYILQGLIASATNLANLPNSLA
                FMIAIDGYVMWLNKRNNTSRKNVYSPDSVEFAFSDENAKYDLPTNIEIVKT
                GEEKIHVGHSGQTTIGTAFSTNPTLAKIKTFYALAVATKYQDSPKKISFPT
                FLFKLMGSKMFLPHTYPPDDPLFGTGVSEVVDILCSNLFIFCGDKKNLVNREDV
                YLGHNAGSYVDPLHMAOLVSGTQVQARNMGSPSQNMILHYOKPPEVDVAMTPPV
                AVWNGNDLADPDVAMLLPKLSNLFHKELIATNHLDFIAMDAPQEVYDEMISM
                AAD"
    CDS
        7..1194
            /codon_start=1
            /product="prelingual lipase"
            /protein_id="CAA00136.1"
            /db_xref="GI:14749"
            /db_xref="SWISS-PROT:P04634"
            /translation="MMLLITVISTFGGAGLFGKLGPNPEANKNISQMTTCYF
            COEEVYTEDGYILGVRIPIHKNNSENIGKRPVYILQGLIASATNLANLPNSLA
            FMIAIDGYVMWLNKRNNTSRKNVYSPDSVEFAFSDENAKYDLPTNIEIVKT
            GEEKIHVGHSGQTTIGTAFSTNPTLAKIKTFYALAVATKYQDSPKKISFPT
            FLFKLMGSKMFLPHTYPPDDPLFGTGVSEVVDILCSNLFIFCGDKKNLVNREDV
            YLGHNAGSYVDPLHMAOLVSGTQVQARNMGSPSQNMILHYOKPPEVDVAMTPPV
            AVWNGNDLADPDVAMLLPKLSNLFHKELIATNHLDFIAMDAPQEVYDEMISM
            AAD"
    alignment_scores:
        Quality: 1166.50      Length: 399
        Ratio: 3.556          Gaps: 1
        Percent Similarity: 82.206      Percent Identity: 54.637
    alignment block:
        US-09-578-063-47 x A01157
    Align seg 1/1 to: A01157 from: 1 to: 1336
16  GluMetTrpLeuLeuIleLeuValAlaTyMetPheGln...ArgAsnVa 31
   ||||| |||||
4  AAGATGTGCTGTATTAATACAAAGTGTGATACAACTTCGAGGTGC 53
   ||||| |||||
31  lAsnSerValHisMetProThrLysAlaValAspProGlnAlaPheMet 48
   ||||| |||||
54  ACATGGCTATATTGAAAACTGGGCTCGAAACCCGGAACCAATATGA 103
   ||||| |||||
48  snIleSerGlnIleIleGlnHisGlnGlyTyTrpCysGlnGlnTyGlu 64
   ||||| |||||
104  ATATTAGTCAGATTAACCTTACTGCGGGAATATTCATGTCAGAAATGA 153
   ||||| |||||
65  ValAlaThrGlnAspGlyTyTrpIleLeuSerValAsnArgIleProArg 81

```

```

154 GTTGTACTGAGATGGCTACTTCTGGGGGCTCTACAGAAATTCCTCATG 203
      |||
81 yLeuValGlnProLysThrGlySerArgProValValLeuLeuGln 98
      |||
204 GAAGATATATCTGAAATATAGGACAGACCTGTGTATTTTCAC 253
      |||
98 lAgLYLeuValGlyAlaSerAsnTrpIleSerAsnLeuProAsn 114
      |||
254 AGGTTTGTATCCATCAGCCACAACTGATTCGAATCTACCAAAAC 303
      |||
115 SerLeuGlyPheIleLeuAlaAspAlaGlyPheAspValTrpMetGly 131
      |||
304 AGCCTGGCTTTATGTCTACAGACGCTGCTATGATGTGGCTGGGGA 353
      |||
131 nSerArgGlyAsnAlaTrpSerArgLysHisLysThrLeuSerIleAsp 148
      |||
354 CAGTGAGGAAATACATGCTCCAGGAAATGATACTCTACACGACT 403
      |||
148 lAspGluPheTrpAlaPheSerTrpAspGluMetAlaArgPheAspLeu 164
      |||
404 CAGTTGAATTCGGCTTACCTTTGATGAATAGCTTAATATGACTT 453
      |||
165 ProAlaValIleAsnPheIleLeuGlnLysThrGlyGlnGlyLysIle 181
      |||
454 CCGGCACAAATAACTCTATGTACGAAACTGACAAAGAGATAC 503
      |||
181 rTyValGlyTySerGlnGlyThrMetGlyPheIleAlaPheSer 198
      |||
504 CTAATGTGTCATCTCCAGGACACCACTGTGTTTCATGCTTTCT 553
      |||
198 hMetProGluLeuAlaGlnLysIleLysMetTyPheAlaLeuAlaPro 214
      |||
554 CCAATCCCTACTGGCCAAATAATCAAGACGTTTATCATAGCTCA 603
      |||
215 lIleAlaThrValLysHisAlaLysSerProGlyThrLysPheLeuLeu 231
      |||
604 GTTGACCGGTGAATACACAAAGTCCCTGAAACATTTCAATTT 653
      |||
231 uProAspMetIleLysGlyLeuPheGlyLysLysGluPheLeuTrp 248
      |||
654 TCCTCATATCTTTCAACCTATGTTGGCAAGAAATGTTCTCC 703
      |||
248 lInThrArgPheLeuArgGlnLeuValIleTyLeuLysGlyGlnValIle 264
      |||
704 ACACCTACTTGTATGATCTTGTGACCAAGTGTGTGACGGAGTT 753
      |||
265 LeuAspGlnIleCysSerAsnIleMetLeuLeuGlyGlyPheAsn 281
      |||
754 CTAGATCTTCTCTGACGAAACATTTATTCATCTGTGATTTGACAA 803
      |||
281 rAsnAsnMetCysMetSerArgAlaSerValTyAlaAlaHisThrLeu 298
      |||
804 GAAAACTTAATATGAGTCGTTTGATGTGTATCTAGGCAATATCCAG 853
      |||
298 lAgLYThrSerValGlnAsnIleLeuHisTrpSerGlnAlaValAsnSer 314
      |||
854 CAGGACATCTGTCAGACTTCTCCACTGGGACAGCTGTATGATCT 903
      |||
315 GlyGluLeuArgAlaPheAspTrpGlySerGluThrLysAsnLeuGly 331
      |||
904 GGGAAATTTCAAGCCTCAACTGGGGAGCCCATCCCAACATGTTTACA 953
      |||
331 sCysAsnGlnProThrProValArgTyArgValArgAspMetThrVal 348
      |||
954 CTACACCGAAGAAACGCTCTCGAATATGATGTGACGCAATGACTGTC 1003
      |||
348 rOThrAlaMetTrpThrGlyGlyGlnAspTrpLeuSerAsnProGluAsp 364
      |||
1004 CAGTTGCAAGTGTGGAACGGTGCATATGATCTGCTGATCCCAAGAT 1053
      |||
365 ValLysMetLeuLeuSerGluValThrAsnLeuIleTyHisLysAsn 381
      |||

```

```

1054 GTGCGATGCTGCTCCCAACTCTCCACCTCTCTTCATTAAGAGAT 1103
      |||
381 eProGluTrpAlaHisValAspPheIleTrpGlyLeuAspAlaProHis 398
      |||
1104 TCTTGCTTCAATACACCTGACTTCTATTCGGCAATGATGCCCTCAAG 1153
      |||
398 rGMetTyAsnGluIleIleHisLeuMetGlnGlnGluLutTrpAsn 413
      |||
1154 AGTTTACATAGAGATTCATGATGTCAGAGAACTATTAAGAT 1200
      |||
seq_name: 'gb_rnLIP
seq_documentation_block:
LOCUS       rnLIP               1355 bp      mRNA           ROD           30-MAR-1995
DEFINITION  Rat mRNA for lingual lipase..
ACCESSION   X02309
VERSION     X02309.1 GI:56595
KEYWORDS    glycoprotein; lipase.
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 1355)
            Doeherty, A.J., Bodmer, M.W., Angal, S., Verger, R., Riviere, C.,
            Lowe, P.A., Lyons, A., Embley, J.S. and Harris, T.J.
            Molecular cloning and nucleotide sequence of rat lingual lipase
            cDNA.
JOURNAL     Nucleic Acids Res. 13 (6), 1891-1903 (1985)
MEDLINE     85215587
COMMENT     Data kindly reviewed (12-FEB-1986) by A.J.P. Doeherty.
FEATURES
            source
                1..1355
                    /organism="Rattus norvegicus"
                    /db_xref="taxon:10116"
                    /product="put. signal peptide (aa -18 to -1)"
                    /protein_id="CAA26179.1"
                    /note="lipase precursor"
                    /codon_start=1
                    /db_xref="GI:56596"
                    /db_xref="SWISS-PROT:P04634"
                    /translation="MWLLITVISTGAGHGLFKLSPGNPEANMNISQNTIYWGTP
                    COEEVVTEDGYILIGYRIIPGKNSNSNIGRPVYVYLOHGLIASATWIANLPNNSLA
                    FMLADAGYDWGLGNSRGNTSRKNTVYSPDSPEWAFSPDMATYDLPATINFTVORT
                    GQEKIHVYVSGSGTIGFIAFTSNPTLAKKIKTFYALPAPVATYDPSPIKISFIP
                    FLKLMGKKKMFPHYTFDEPLGTEGNSREVLIDLCNTLIFGCGKKKNLNSRPFV
                    YLGHNPAGTIVODPLHMAOLVRSGRFOAFNMGSSQNMKLNOKTPPEIDYSAATFV
                    AVMWGNDIADPDQVAMLLPKLSNLFHKEILVYVNHLDFTWADADQEVYENIMISM
                    AED"
            mat_peptide
                103..1233
                    /product="lipase (aa 1-377)"
            misc_feature
                505..507
                    /note="aa 153 pot. lipid binding site"
            misc_feature
                1355
                    /note="polya site"
BASE COUNT  400 a      291 c      276 g      387 t      1 others
ORIGIN
alignment_scores:
    Quality: 1166.50      Length: 399
    Ratio: 3.556          Gaps: 1
    Percent Similarity: 82.206      Percent Identity: 54.637
alignment_block:
US-09-578-063-47 x rnLIP
Align seg 1/1 to: rnLIP from: 1 to: 1355
16 GluMetTrpLeuLeuIleLeuValAlaTyMetPheGln...ArgAsnVa 31
      |||

```

331 scysansgInProthProvalaIgTygArValaArgasPmetInValp 348  
 996 CTACAGACCAAGAAACGCTCTGTAATATCATATGTGCAGCCATGACTGTGC 1045  
 348 rOthAlaMetPrThrGlyGlyGlnAspPrLysSerAsnProGluAar 364  
 1046 CAGTTGGCACTGTGGAACGGTGGCATGACATCTGTGGCTATGCCCAAGAT 1095  
 365 ValLysMetLeuLeuSerGluValThrAsnLeuIleTyrlHisLysAsn11 381  
 1146 TCTTGCCCTACAACTCACTGGACTTATCTGGCAATGATGCCCTCCCAAG 1195  
 398 rGMetTyAsnGlnIleIleHisLeuMetGlnGlnGluIleLurThrasn 413  
 1196 AGGTTTCATATGATGATTTCCATGATGACGACAAAGACTAAAGAAAT 1242  
 seq\_name: gb\_om:CFLIPASE  
 seq\_documentation\_block: 1651 bp mRNA MAM 18-JUN-1998  
 LOCUS CFLIPASE  
 DEFINITION Canis familiaris mRNA for lipase.  
 ACCESSION Y13899  
 VERSION Y13899.1 GI:2204112  
 KEYWORDS DGL gene; gastric lipase; triacylglycerol lipase.  
 SOURCE dog.  
 ORGANISM Canis familiaris  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 1 (bases 1 to 1651)  
 Vaganay,S., Jolliff,G., Bertaux,O., Toselli,E., Devignes,M.D. and  
 Benicourt,C.  
 The complete cDNA sequence encoding dog gastric lipase  
 DNABase Seq. 6 (4), 257-262 (1998)  
 2 (bases 1 to 1651)  
 Benicourt,C.  
 Direct Submission  
 Submitted (10-JUN-1997) C. Benicourt, Ecole Normale Supérieure de  
 Cachan, L.I.R.B.A (Laboratoire Interdisciplinaire de Recherche en  
 Biologie Appliquée), 61 avenue du Président Wilson, F-94235 Cachan  
 Cedex, FRANCE  
 FEATURES  
 source  
 1. 1651  
 /organism="Canis familiaris"  
 /db\_xref="taxon:9615"  
 /dev\_stage="adult"  
 /tissue\_type="stomach mucosa"  
 54. .1250  
 /gene="DGL"  
 54. .1250  
 /gene="DGL"  
 /EC\_number="3.1.1.3"  
 /codon\_start=1  
 /product="triacylglycerol lipase"  
 /protein\_id="CAA74198.1"  
 /db\_xref="GI:2204113"  
 /db\_xref="SWISS-PROT:P80035"  
 /translation="MWLLPAAVSTLGLTTHGLFGRKLHPTNPEVTNNISOMITTYGQY  
 PAEYEVATDEGQYILGIDRIPIYGRKNSENIGRFAVALQHGILASATNWSINLPNSGL  
 AFLIADGAYVWVGNSGRNWMARNLVYSPDSVEFMAFSPDEMAKYDLPATIDILKKK  
 TGQDKLHYVGHSSQCTTGFTAFSPNPLARKTFVALAPATVRYETILNKIMLIVP  
 SFELKIFGNKILTFPHFDFDLATEVCSREYDILCSNALFTICGFDTNMLNKRSD  
 VYLSHNPAQTSVQVNLHWSQAVKSGKQADWGSPPVQNMHHYHQSMPYVNLDMHVP  
 IAVNGNDLADPHDHDVLDLISKLPNLIYHRKIPRYNHLDFIWMADAPQAVYNEIVSM  
 MGTGNK"  
 54. .110  
 /gene="DGL"  
 /evidence="experimental"

mat\_peptide 111..1247  
 /gene="DGL"  
 polyA\_signal 1602..1607  
 BASE COUNT 432 a 419 c 352 g 448 t  
 ORIGIN

alignment\_scores:  
 Quality: 1154.00 Length: 398  
 Ratio: 3.486 Gaps: 5  
 Percent similarity: 83.166 Percent identity: 54.271

alignment\_block:  
 US-09-578-063-47 x CELIPASE ..

Align seg 1/1 to: CELIPASE from: 1 to: 1651

```

17 MetTrpLeuLeuIleValAlaIleValMetPheGlnArgAsnValAsn 33
|||||.....:|||||.....:|||||.....:
54 ATGTGGCTGCTATTAAACAGCGGCGAGTGTATCT...ACACTGGAAAC 100
33 rValHis.....:.....:MetProThrLysAlaValAspProG 44
:|||||.....:|||||.....:
101 CACACATGGTTATTGGAAATATACATCCCA.....AACCTG 141
44 LuAlaPheMetAsnIleSerGluIleIleGlnHisGlnGlyTyrProCys 60
|||||.....:|||||.....:
142 AAGTACACATGAATATAGTACAGATGATCCTACTGGGATACCAGCT 191
61 GluGluTyrGluValAlaIleThrGlnAspGlyTyrIleLeuSerValAsn 77
|||||.....:|||||.....:
192 GAGGAAATAGAGTGTGACCGAAGACGGTATATCTTGGGATCGACAG 241
77 GlieProArgGlyLeuValAlaProLysLysThrGlySerArgProVal 94
|||||.....:|||||.....:
242 AATTCCTTATGGGAGGAAAAATTCAGAGATATAGCCCGAGACCTGTTG 291
94 aLeuLeuGlnHisLysLeuValGlyGlyAlaSerAsnTrpIleSerAsn 110
|||||.....:|||||.....:
292 CATTTTGGACACCGTTTGGCTGCATCAGCCCAAACTGATCTCCAAC 341
111 LeuProAsnAsnSerLeuGlyPheIleLeuAlaAspAlaGlyPheAsp 127
|||||.....:|||||.....:
342 CTGCCCAACACAGCTGGCTTCATCCGCGGAGCGGGGTACGAGCT 391
127 LTrpMetGlyAsnSerArgLysAlaIleTrpSerArgLysHisLysThr 144
|||||.....:|||||.....:
392 GTGGCTGGGAGACAGAGGCGCAACACCTGGCGGAGAGATCTGTACT 441
144 euSerIleAspGlnAspGluPheTrpAlaPheSerTyrAspGluMetAla 160
|||||.....:|||||.....:
442 ACTCCCGGACACTCGCATATCTGGCTTTCAGCTTTCACAGATGCT 491
161 ArgPheAspLeuProAlaValIleAsnPheIleGlnLysThrGlyGly 177
|||||.....:|||||.....:
492 AATATGACCTTCGCGCCACATGACTCATCTTGAGAAAAACGGGAGA 541
177 nGluLysIleTyrValGlyTyrSerGlnGlyThrMetGlyPheI 194
|||||.....:|||||.....:
542 GGACACAGACACTAGTGGCCATTCACAGGCAACCAATGATGTTTCA 591
194 LeAlaPheSerThrMetProGluLeuAlaGlnLysIleLysMetTyrPhe 210
|||||.....:|||||.....:
592 TCGCCTTTTCCACCAATCCAGCTGCGGAAACGATCAAACTCTTAT 641
211 AlaLeuAlaProIleAlaThrValLysHisAlaLysSerProGlyThr 227
|||||.....:|||||.....:
642 GCATTAGCTCCCGTGGCCACCGTGAATACACGAAACCTTTAAACA 691
227 sPheLeuLeuLeuProAspMetMetIleLysGlyLeuPheGlyLysG 244
|||||.....:|||||.....:
692 ACTCATGCTGCTGCTGCTCTCAAGCTTATATTGGAACAAAAA 741

```

```

244 LuPheLeuTyrGlnThrArgPheLeuArgGln...LeuValIleTyrLeu 259
|||||.....:|||||.....:
742 TATTC...TACCACACACACTTCTTGATCAATTTCTGCCACCGAGTA 788
260 CysGlyGlnValIleLeuAspGlnIleCysSerAsnIleMetLeuLeu 276
|||||.....:|||||.....:
789 TGCCTCCCGAGAGCGGTGATCTCTTCGACACACAGCGCTTATCAT 838
276 uGlyGlyPheAsnThrAsnAsnMetAsnMetSerArgAlaSerValTyr 293
|||||.....:|||||.....:
839 TTGTGATTTGCACTATGAACCTTGAACATGATGCTGGTGGTGTATC 888
293 LeAlaHisThrLeuAlaGlyThrSerValGlnAsnIleLeuHisTrpSer 309
|||||.....:|||||.....:
889 TGTCAATTAATCCAGACGAAACATCGTTTCAGAACGTGCTCCACTGTTC 938
310 GlnAlaValAsnSerGlyGluLeuArgAlaPheAspTrpGlySerGluThr 326
|||||.....:|||||.....:
939 CAGCGCTTAAAGTCTGGGAAGTCCACAGCTTTGACTGGGAGCCAGT 988
326 rLysAsnLeuGluLysCysAsnGlnProThrProValArgTyrArgVala 343
|||||.....:|||||.....:
989 TCAGAAACATGATGCACTATCATCAGACATGCCCTCTACTACAACTGA 1038
343 rGAspMetThrValProThrAlaMetTrpThrGlyGlnAspTrpLeu 359
|||||.....:|||||.....:
1039 CAGACATCATGCTGCCAATCGCAGTGTGAGAGCGGAGCAACACTTGCTG 1088
360 SerAsnProGluAspValLysMetLeuLeuSerGlnValAlaThrLeu 376
|||||.....:|||||.....:
1089 GCCGACCTCCAGATGTGACCTTGTCTTCCAGCTCCCAAACTCAT 1138
376 eTyrHisLysAsnIleProGluTrpAlaHisValAspPheIleTrpGly 393
|||||.....:|||||.....:
1139 TTACCACAGGAAGATTCTCCTTACATCAGCTTGACCTTATCTGGGCA 1188
393 euAspAlaProHisArgMetTyrAsnGluIleIleHisLeuMet 407
|||||.....:|||||.....:
1189 TGGATGCCCTCCACAGCGTTTCAATGAATGTTTCCATGATG 1232

```

seq\_name: gb\_pat:AR039022

seq\_documentation\_block: 1137 bp DNA PAT 29-SEP-1999  
 LOCUS AR039022  
 DEFINITION Sequence 2 from patent US 5807726.  
 ACCESSION AR039022  
 VERSION AR039022.1 GI:5958385  
 KEYWORDS

SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 1137)  
 AUTHORS Blanchard,C., Benicourt,C. and Junien,J.  
 TITLE Nucleic acids encoding dog gastric lipase and their use for the  
 JOURNAL production of polypeptides  
 PATENT: US 5807726-A 2,15-SEP-1998;  
 FEATUERS Location/Qualifiers  
 source 1..1137

BASE COUNT 296 a 308 c 249 g 284 t  
 ORIGIN

alignment\_scores:  
 Quality: 1141.50 Length: 372  
 Ratio: 3.624 Gaps: 3  
 Percent similarity: 84.677 Percent identity: 56.452

alignment\_block:  
 US-09-578-063-47 x AR039022 ..  
 Align seg 1/1 to: AR039022 from: 1 to: 1137

```

37  ProthrIysAlaValAspProGluAlaPheMetAsnIleSerGluIleI 53
    ||||| ..... ||||| ||||| |||||
19  CCCACA.....AACCTGAAGTGCACATGATATTAAGTCAGATGAT 59
53  eGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaIaThrGluAsp 70
    ||| ..... ||||| ||||| ||||| |||||
60  CACCTACTGGGGATACCCAGCTGAGGAAATAGAAAGTTGTGACCGAAGACG 109
70  IyTyrIleLeuSerValAsnArgIleProArgGlyLeuValGlnProLys 86
    ||||| ..... ||||| ||||| ||||| |||||
110 GTTATATCTCTGGGATGCACAGAAATCTTATGGAGGAAAAAATTCAGAG 159
87  LysThrGlySerArgProValIleLeuGlnIleHisGlyLeuValGlyI 103
    ||| ..... ||||| ||||| ||||| |||||
160 AATATAGCGCGAGACCTGTTGCACTTTTGCACACGCGTTGCTGCATC 209
103 yAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleI 120
    ||||| ..... ||||| ||||| ||||| |||||
210 AGCCACAACCTGATCTCCAACTGCCCCAACACAGCCTGGCCTTCATCC 259
120 euaIaAspAlaGlyPheAspValIleTrpMetGlyAsnSerArgGlyAsnAla 136
    ||||| ..... ||||| ||||| ||||| |||||
260 TGCGGCAGCGCGGTACGACCTGTGGCTGGGGAAGCAGGAGGGAACACC 309
137 TrpSerArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAl 153
    ||||| ..... ||||| ||||| ||||| |||||
310 TGCGCCAGGAGAACTGTACTACTCGCCGACCTCGCTGGAATCTGGGC 359
153 aPheSerTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnP 170
    ||||| ..... ||||| ||||| ||||| |||||
360 TTATAGCTTTGACGATGCTTAATATAGACCTTCGCCACCATGTGACT 409
170 heIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSer 186
    ||||| ..... ||||| ||||| ||||| |||||
410 TCACCTTGAAGAAACGGGAGACGACAAAGCTACACTAGCTGGCCATCC 459
187 GlnGlyThrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAl 203
    ||||| ..... ||||| ||||| ||||| |||||
460 CAGGCGACACCATGTGTTTCATCGCCTTTCACCAATCCCAAGCTGGC 509
203 aGlnLysIleLysMetTyrPheAlaIleuAlaProIleAlaThrValLysH 220
    ||||| ..... ||||| ||||| ||||| |||||
510 GAACGGGATCAAAACCTTATGCAATAGCTCCGCTGCCACCGGAAGT 559
220 IAlaLysSerProGlyThrLysPheLeuLeuLeuProAspMetMetIle 236
    ||||| ..... ||||| ||||| ||||| |||||
560 ACACCGAAACCTGTAAACAACTCATGCTGCTCCCTTCCTCTTC 609
237 LysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArgPheLeuAr 253
    ||| ..... ||||| ||||| ||||| |||||
610 AAGCTTATATTTGGAACAAATATTC...TACCACACACACTCTTTGA 656
253 gGln...LeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleC 269
    ||| ..... ||||| ||||| ||||| |||||
657 TCAATTTCTCGCCACCGAGTATGCTCCCGGAGACGGTGGATCTCTCT 706
269 ySerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsn 285
    ||||| ..... ||||| ||||| ||||| |||||
707 GCACCAACGCCCTGTTATCATTTGTGTGATTTGACACATATGACTTGAAC 756
286 MetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVa 302
    ||||| ..... ||||| ||||| ||||| |||||
757 ATGAGTCGCTTGATGTATCTCATATTCACAGCAGACGAGAACATCGGT 806
302 IGlAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuAsnArg 319
    ||||| ..... ||||| ||||| ||||| |||||
807 TCAGAAGCTGCTCCACTGCTCCAGGCTGTAAGCTGGGAAGTCCACAG 856
319 IaPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnPro 335
    ||||| ..... ||||| ||||| ||||| |||||
857 CTTTGTGACTGGGAAGCCCACTTCAGAACATGATGACACTATCATCAGAGC 906
336 ThrProValArgTyrArgValAlaArgAspMetThrValProThrAlaMetTr 352

```

```

907 ATGGCTCCCTACTACAACTGACAGACATGCATGTGCCAATCGAGGTG 956
352 pThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuI 369
    ||||| ..... ||||| ||||| ||||| |||||
957 GAACGTGGCAACGACTGTGGCGGACCCYACAGATGTGACCTTTGTC 1006
369 euserGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAla 385
    ||||| ..... ||||| ||||| ||||| |||||
1007 TTTCCAACCTCCCAATCTCAATTAACACAGGAAGATTCTCTTACAAT 1056
386 HisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnG 402
    ||||| ..... ||||| ||||| ||||| |||||
1057 CACTTGGACTTATCTGGCCATGATGCCCTCAAGCGTTTCAATGA 1106
402 uIleIleHisLeuMet 407
    ||||| ..... ||||| ||||| ||||| |||||
1107 AATGTTCATGATG 1122

seq_name: gb-pat:AR092633

seq_documentation_block:
LOCUS AR092633 1137 bp DNA PAT 08-SEP-2000
DEFINITION Sequence 2 from patent US 5998189.
ACCESSION AR092633
VERSION AR092633.1 GI:10019385
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1137)
AUTHORS Blanchard,C., Benicourt,C. and Junten,J.
TITL E Polypeptide derivatives of dog gastric lipase and pharmaceutical
JOURNAL Patent: US 5998189-A 2 07-DEC-1999;
FEATURES
Location/Qualifiers
source 1..1137
BASE COUNT 296 a 308 c 249 g 284 t
ORIGIN

alignment_scores:
Quality: 1141.50 Length: 372
Ratio: 3.624 Gaps: 3
Percent Similarity: 84.677 Percent Identity: 56.452

alignment_block:
US-09-578-063-47 x AR092633 ..

Align seg 1/1 to: AR092633 from: 1 to: 1137

37 ProthrIysAlaValAspProGluAlaPheMetAsnIleSerGluIleI 53
    ||||| ..... ||||| ||||| ||||| |||||
19 CCCACA.....AACCTGAAGTGCACATGATATTAAGTCAGATGAT 59
53 eGlnHisGlnGlyTyrProCysGluGluTyrGluValAlaIaThrGluAsp 70
    ||| ..... ||||| ||||| ||||| |||||
60 CACCTACTGGGGATACCCAGCTGAGGAAATAGAAAGTTGTGACCGAAGACG 109
70 IyTyrIleLeuSerValAsnArgIleProArgGlyLeuValGlnProLys 86
    ||||| ..... ||||| ||||| ||||| |||||
110 GTTATATCTCTGGGATGCACAGAAATCTTATGGAGGAAAAAATTCAGAG 159
87 LysThrGlySerArgProValIleLeuGlnIleHisGlyLeuValGlyI 103
    ||| ..... ||||| ||||| ||||| |||||
160 AATATAGCGCGAGACCTGTTGCACTTTTGCACACGCGTTGCTGCATC 209
103 yAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleI 120
    ||||| ..... ||||| ||||| ||||| |||||
210 AGCCACAACCTGATCTCCAACTGCCCCAACACAGCCTGGCCTTCATCC 259
120 euaIaAspAlaGlyPheAspValIleTrpMetGlyAsnSerArgGlyAsnAla 136

```

```

|||||
260 TGGCCAGCCCGGGTACGACGTGTGGCTGGGAAACGCGGGGCAACACC 309
137 TTPSerArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAl 153
|||||
310 TGGGCGCAGAGAGATCTGACTACTGCGCCGACCTCCGCGAATTCGGGGC 359
153 aPheSerTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsp 170
|||||
360 TTTCAGCTTTGACAGATGCTGCTAAATATGACCTCCCGCCACCATGTGACT 409
170 heIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSer 186
|||||
410 TCATCTTGAAGAAACGGACAGGACAACTACACTACCTGCGCCCAATTC 459
187 GlnGlyThrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAl 203
|||||
460 CAGGGCACCACCATTTGGTTTCATCGCTTTTCACCAATCCCAAGCTGGC 509
203 acGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysH 220
|||||
510 GAAACGATCAAAACCTCTATGCAATTAGCTCCCGTTGCCACCGTGAAGT 559
220 IAspAlaLysSerProGlyThrLysPheLeuLeuLeuProAspMetMetIle 236
|||||
560 ACACCGAAACCGCTGTTAAACAACTCATGCTGCTCCCTCGCTCTCTTC 609
237 LysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArgPheLeuAl 253
|||||
610 AACCTTATATTGGAAACAAATATTC...TACCCACACCACTTCTTTGA 656
253 gGln...LeuValIleTyrLeuGlyGlnValIleLeuAspGlnIleC 269
|||||
657 TCATATTCTCGCCACCGAGGTATGCTCCGCGACAGGTGATCTCTCT 706
269 ySerSerIleMetLeuLeuLeuGlyGlyPheAsnThrAsnMetAsn 285
|||||
707 GCGACAGCCCGCTGTTATCATTTGATTTGACACTATGAACCTTGAAAC 756
286 MetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVa 302
|||||
757 ATAGTCGGCTGGATGTGATATCTGTCAATATCCACAGAGAACATCGGT 806
302 IGlinaGlnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArg 319
|||||
807 TCGAACGCTGCTCCACTGCTCCAGGCTGTAACTCTGGGAAGTTCAG 856
319 IapheAspTrpGlySerGluThrLysAsnLeuLysCysAsnGlnPro 335
|||||
857 CTTTGACTGGGGAGCCCGCATTCAGAACATGATGCACTATCATCAGAGC 906
336 ThrProValArgTyrArgValaArgAspMetThrValProThrAlaMet 352
|||||
907 ATGCCCTCCCTACTACAACTGACAGACATGCAATGCCAATGCCAGTGT 956
352 rThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeu 369
|||||
957 GAACGCTGGGAAAGCACTGCTGCGCCAGCCCTCAGATGTGACCTTTTGC 1006
369 euSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAla 385
|||||
1007 TTTCAGAGCTCCCAATCTCATTTACCAAGGAAGATTCCTCTTCAAT 1056
386 HisValAspPheIleTyrPglLysAspAlaProHisArgMetTyrAsnG 402
|||||
1057 CACTTGGAGCTTATCTGGGCGCATGATGCCCTCAAGCGTTTACATGA 1106
402 uIleIleHisLeuMet 407
|||||
1107 AATTGTTCCATGATG 1122
seq_name: gp_pat:AR039023

```

```

seq_documentation_block: 1140 bp DNA PAT 29-SEP-1999
LOCUS AR039023 4 from patent US 5807726.
DEFINITION Sequence
ACCESSION AR039023
VERSION AR039023.1 GI:5958386
KEYWORDS
SOURCE
ORGANISM
Unknown.
REFERENCE
1 (bases 1 to 1140)
AUTHORS
Blanchard,C., Benicourt,C. and Julien,J.
TITLE
Nucleic acids encoding dog gastric lipase and their use for the
production of polypeptides
JOURNAL
Patent: US 5807726-A 4 15-SEP-1998;
FEATURES
location/Qualifiers
source
1..1140
BASE COUNT 297 a 308 c 250 g 285 t
ORIGIN
alignment_scores:
Quality: 1141.50 Length: 372
Ratio: 3.624 Gaps: 3
Percent Similarity: 84.677 Percent Identity: 56.452
alignment_block:
US-09-578-063-47 x AR039023
Align seg 1/1 to: AR039023 from: 1 to: 1140
37 ProThrLysAlaValAspProGluAlaPheMetAsnIleSerGluIleI 53
|||||
22 CCCACA.....AACCTGAAAGTACCACGAATATATGATGATGAT 62
53 ecGlnHisGlnGlyTyrProGlyGlnGluLysValAlaThrGluAsp 70
|||||
63 CACCTACTGGGATCCAGCTGAGGAAATATGAAATGTTGACCAGGAGAG 112
70 LYTrrIleLeuSerValAsnArgIleProArgGlyLeuValGlnProLys 86
|||||
113 GTTATATCCTTGGGATCGACAGATTCCTTATGGGAGGAAATTCAGAG 162
87 LysThrGlySerArgProValValIleLeuGlnHisGlyLeuValGly 103
|||||
163 AATATAGCGCGGAGACCTGTGATTTTTCACACAGGTTTGTCTGCATC 212
103 yAlaSerAsnTrpIleSerAsnLeuProAsnSerLeuGlyPheIleL 120
|||||
213 AGCCACAAACTGGATCTCCAACTGCGCAACACACAGCTGGCCTTATCC 262
120 euAlaAspAlaGlyPheAspValTrpMetGlyAsnSerArgLysAsnAla 136
|||||
263 TGGCGCAGCCCGGGTACGACGTGTGGCTGGGAAACGAGGGGCAACACC 312
137 TTPSerArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAl 153
|||||
313 TGGGCGCAGAGAGATCTGACTACTGCGCCGACCTCCGTGGAATTCGGGC 362
153 aPheSerTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsp 170
|||||
363 TTTCAGCTTTGACAGATGCTGCTAAATATGACCTCCCGCCACCATGTGACT 412
170 heIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSer 186
|||||
413 TCATCTTGAAGAAACGGACAGGACAACTACACTACACTGTGGCCATTC 462
187 GlnGlyThrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAl 203
|||||
463 CAGGGCACCACCATTTGGTTTCATCGCTTTTCACCAATCCCAAGCTGGC 512
203 acGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysH 220
|||||

```

```
513 GAAACGGATCAAAACCTTATGCAATAGTCCCGTTGCCACCGTAGT 562
220 IahAlaLysSerProGlyThrLysPheLeuLeuLeuProAspMetMetIle 236
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
563 ACACCGAAACCCGTGTAACAAACATCATGCTGCCCTTGTCTCTCTC 612
237 LysGlyLeuPheGlyLysGlyLysGlyPheLeuTyGlnThrArgPheLeuAr 253
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
613 AACCTATATTTGGAAACAAATATTC...TACCACACCCACTTCTTGA 659
253 gGln...LeuValIleTyLeuCysGlyGlnValIleLeuAspGlnIleC 269
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
660 TCAATTTCTGCCACCGAGGTATGTCGCCGAGAGGTGATCTCTCT 709
269 ySerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnMetAsn 285
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
710 GCAGCAACGCGCGTTTATCATTTGTGATGTGACACTATGAACCTGAAAC 759
286 MetSerArgAlaSerValTyAlaAlaIleAsnThrLeuAlaGlyThrSerVa 302
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
760 ATGAGTGGCTGGATGTGTGTCTGCACATTAATCCAGAGAAACATCGGT 809
302 IGIAsnIleLeuIleAsnIleTrpSerGlnAlaValAsnSerGlyLeuLeuArg 319
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
810 TCAGAACGTGCTCCACTGCTGCCAGGCTGTTAAGTGTGGGAGATTCGAC 859
319 IAPheAspTrpGlySerGlyThrLysAsnLeuGlnLysCysAsnGlnPro 335
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
860 CTTTGTACTGGGGAACCCGAGTTCAGAAACATGATGCACATCATCAGAGC 909
336 ThrProValArgTyArgValAlaArgAspMetThrValProThrAlaMetTr 352
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
910 ATGCTCTCCCTACTACAACTGACAGACATGATGCGCAATCGCAGAGTG 959
352 pThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuL 369
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
960 GAACGGTGGCAACGACTGCTGCGCCGACCTCAGATGTTGACCTTTGC 1009
369 euSerGluValThrAsnLeuIleTyThrLysAsnIleProGluTrpAla 385
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
1010 TTCCCAAGCTCCCAATCTCATTTACCAAGAAAGATTCTCTTACAAT 1059
386 HisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyArgnG 402
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
1060 CACTTGACTTTATCTGGGCAATGATGCCCTCAAGCGGTTTACAATGA 1109
402 uIleIleHisLeuMet 407
      ::::::::::: :::::::::::
1110 AATTGTTTCCATGATG 1125
```





```

1 MetLeuGlnThrLeuSerArgGlnTrpIleValSerHisArgMetGlnLeu 17
1 ATGTGGAAACCTTGTCAGAACAGTGGATGTGTCACACAGATGGAAAT 50
17 tTrrLeuLeuIleLeuValAlaIleTyrMetPheGlnArgAsnValAsnSerV 34
51 GTGGCTTCTGATCTGTGGCGCTATATGTCACAGAAATGGAATTTGAG 100
34 aLHisMetProThrLysAlaValAlaValAspProGlnAlaPheMetAsnIleSer 50
101 TACATATGCCAAGCTAAAGCTGTGGACCCAGAAAGCAATTCATGAATTTAGT 150
51 GluIleIleGlnHisGlnGlyTyrProCysGlnGluIleTyrGlnValAlaThr 67
151 GAAATCATCCAAACATCAAGGCTATCCCTGTGAGGAATATGAAGTGGCAAC 200
67 rGluAspGlyTyrIleLeuSerValAsnArgIleProArgGlyLeuValG 84
201 TGAAGATGGTATATCCTTTCTGTAAACAGGATTCCTCGAGGCCCTAGTGC 250
84 InProLysLysThrGlySerArgProValValIleLeuGlnHisGlyLeu 100
251 AACCTAAAGACAGAGTTCACAGGCTGTGGTGTACTGCACAGCATGGCTTA 300
101 ValGlyGlyAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGln 117
301 GTTGAGGTGGTACCACTGATTTCCAAACCTGCCCAACAATAGCTTGGG 350
117 yPheIleLeuAlaAspAlaGlyPheAspValIleTyrMetGlyAsnSerArgG 134
351 CTTCACTTGGCAGATGCTGTTTGGACGTGGATGGAGGAAACACAGAGG 400
134 LysAsnAlaTrpSerArgLysHisLysThrLeuSerIleAspGlnAspGln 150
401 GAAACGCTGGCTCGAAACACCAAGACACTCTCCATAGACCACAAAGATAG 450
151 PheTrpAlaPheSerTyrAspGlnMetAlaArgPheAspLeuProAlaVal 167
451 TTCGGGCTTTCAGTTATGATGAGATGGCTAGTGTGACCTTCCGCAAGT 500
167 ILeAsnPheIleLeuGlnLysThrGlyGlnGlyLysIleTyrValG 184
501 GATAAACTTATTTTTCAGAAACCGGCGCAGAAAGATCTATATATGTGCG 550
184 LYTyrSerGlnGlyThrThrMetGlyPheIleAlaPheSerThrMetPro 200
551 GCTATTCACAGGGCACACCAATGGCTTTATTTGCAATTTCCACCATGTGCCA 600
201 GluLeuAlaGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaThr 217
601 GAGCTGGCTCAGAAATCAAAATGATTTTGTCTTAGCACCATTAGCCAC 650
217 rValLysHisAlaLysSerProGlyTyrLysPheLeuLeuLeuProAspM 234
651 TGTTPAAGCATCAAAAACCCCGGACCAAAATTTTGTGCTGCCAGATA 700
234 eTMeIleLysGlyLeuPheGlyLysLysGlnPheLeuTyrGlnThrArg 250
701 TGATGATCAAGGATGTTTGGCAAAAAGAAATTTCTATACAGACCCAGA 750
251 PheLeuArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAspGln 267
751 TTTTCACACAACTGTATATTACCTTGTGGCCAGGTGATTCCTTGATCA 800
267 nIleCysSerAsnIleMetLeuLeuGlyGlyPheAsnThrAsnAsnM 284
801 GATTGTGTATATATCATGTTACTTCTGTGGTGAATCAACACCAACATA 850
284 eTanMeSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThr 300
851 TGAACATGAGCCGAGCAAGTGTATATGCTGCCACACACTTGTGGAAACA 900

```

```

301 SerValGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGlyLeu 317
901 TCTGTGCAAAATATCTACACTGAGCGACGACGAGTAATCTGAGTA 950
317 uArgAlaPheAspTrpGlySerGlnThrLysAsnLeuGlnLysCysAsnG 334
951 CCGGCAATTTGACTGGGGAGTGAACCAAAAATCTGAAAAAATGCCAATC 1000
334 InProThrProValArgTyrArgValArgAspMetThrValProThrAla 350
1001 AGCCACTCTCTGTAGGTACAGAGTCAGAGATATGACCGTCCCTACACA 1050
351 MetTrpThrGlyGlnAspTrpLeuSerAsnProGlnAspValLysMe 367
1051 ATGTGACAGAGAGGTACAGACTGGCTTCAATCCAGAAAGCTGAAAT 1100
367 tLeuLeuSerGlnValThrAsnLeuIleTyrHisLysAsnIleProGlnT 384
1101 GCTGCTCTGAGGTGACCAACCTCATCTACATTAAGAATATTCCTGAAT 1150
384 rPalAlaHisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyr 400
1151 GGGCTCACGTGATTTCACTGGGGTGTGAGTCTCTCACCGTATGTAC 1200
401 AsnGluIleIleHisLeuMetGlnGlnGlnIleThrAsnLeuSerGlnGln 417
1201 AATGAATCATCCATCTGATGAGCAGAGAGACCAACCTTCCACAGG 1250
417 yArgCysGlnAlaValLeu 423
1251 ACGGTGTGAGGCCGTATATG 1269

```

seq\_name: /SID58/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF45131

seq\_documentation\_block:

ID AAF45131 standard; cDNA; 2044 BP.

XX AAF45131;

XX 30-MAR-2001 (first entry)

XX

XX Human TANGO 294 cDNA.

XX

KM Gene therapy: TANGO protein; INTERCEPT protein; neurological disorder;

KM central nervous system; focal brain disorder; bipolar affective disorder;

KM global diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;

KM senile dementia; Huntington's disease; amyotrophic lateral sclerosis;

KM Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;

KM neuropsychiatric; psychoactive substance use; anxiety; ss.

OS Homo sapiens.

PN WO200077239-A2.

XX 21-DEC-2000.

PD

XX 24-MAY-2000; 2000MO-US14858.

PF

XX 14-JUN-1999; 99US-0333159.

PR

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Claim 1; Fig 6; 359pp; English.

XX The present invention relates to TANGO or INTERCEPT proteins and coding  
 CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,  
 CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding  
 CC sequences are useful for the treatment of neurological disorders such as  
 CC central nervous system (CNS) disorders, CNS-related disorders, focal  
 CC brain disorders, global-diffuse cerebral disorders and other  
 CC neurological and cerebrovascular disorders. The CNS disorders include  
 CC Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic  
 CC lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,  
 CC autonomic function disorders such as hypertension and sleep disorders;  
 CC neuropsychiatric disorders, psychosocial substance use disorders,  
 CC anxiety, and bipolar affective disorder.

XX Sequence 2044 BP, 618 A, 401 C, 460 G, 565 T, 0 other;

alignment\_scores:  
 Quality: 2247.00 Length: 423  
 Ratio: 5.312 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 us-09-578-063-47 x AAF45131

Align seg 1/1 to: AAF45131 from: 1 to: 2044

1 MetLeuGluThrLeuSerArgInPrIleValSerHisArgMetGlu 17  
 126 ATGTTGGAACCTGTCACAGACGTGATGTCACACAGAAATGAAT 175  
 17 TTPRLeuLeuIleuValAlaTyrMetPheGlnArgAsnValAsnSer 34  
 176 GTGGCTTCTGATCTGGTGGGTATGATTCACAGAAATGTGAATCAG 225  
 34 AlHisMetProThrLysAlaValAspProGlnIlePheMetAsnIleSer 50  
 226 TACATAGCCACTAAAGCTGTGAGCCAGAGCAATTCATGAAATTAAT 275  
 51 GluIleIleGlnHisGlnGlyTyrProCysGluGluTyrGluValAla 67  
 276 GAATTCATCCAAATCAAGGCTATCCCTGTGAGAAATATGAAGTCGCAAC 325  
 67 rGluAspGlyTyrIleLeuSerValAsnArgIleProArgGlyLeuVal 84  
 326 TGAAGATGGGTATCTCTTCTGTTAAAGATTCCTGAGGCTAGTGC 375  
 84 InPrLysLysThrGlySerArgProValIleLeuGlnHisGlyLeu 100  
 101 ValGlyLysAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeu 117  
 426 GTTGAGAGGTGCTAGCAACTGGATTCCCACTGCCCAACATATAGCTGG 475  
 117 yPheIleLeuAlaAspAlaGlyPheAspValTyrMetGlyAsnSerArg 134  
 476 CTTCATCTGCGACATGCTGTTTGTGACGTGTGATGGGAAACAGCAGG 525  
 134 LysAlaAlaTyrSerArgLysHisLysThrLeuSerIleAspGlnAsp 150  
 526 GAAACGCTGCTCGAAACACAGACACTCTCATATGACCAAGATGAG 575  
 151 PheTrpAlaPheSerTyrAspGluMetAlaArgPheAspLeuProAla 167  
 576 TTCGGGCTTTCAGTATGATGAGATGGCTATGCTTCCCTGCAGT 625  
 167 LLeAsnPheIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrVal 184  
 626 GATTAACCTTATTTTGCAGAAACGGCCAGGAAAGATCTATATATGTCG 675  
 184 LyrSerGlnGlyThrThrMetGlyPheIleAlaPheSerThrMetPro 200

676 GCTATTCACAGGCGACCATGGGCTTATTTGCATTTCCACCATGCA 725  
 201 GluLeuAlaGlnLysIleLysMetTyrPheAlaLeuAlaProIleAla 217  
 726 GAGCTGGCTCAGAAATCAAAATGATTTCTTTAGCACCATAGCCAC 775  
 217 rValLysHisAlaLysSerProGlyThrLysPheLeuLeuProAsp 234  
 776 TGTTAAGCATCAAAAGCCCGGACCAAAATTTTGTGGTGCAGATA 825  
 234 eMetIleLysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArg 250  
 826 TGATGATCAAGCATTTGTTGGCAAAAAGATTTCTTATACACACCA 875  
 251 PheLeuArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAsp 267  
 876 TTTCACAGCAACTGTTATTTACCTTTGTGGCAGGTGATTCATCA 925  
 267 nIleCysSerAsnIleMetLeuLeuGlyGlyPheAsnThrAsnAsn 284  
 926 GATTTGATGATATATCATGTTACTTCTGGGTGATTCACACCAACA 975  
 284 eLAsnMetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThr 300  
 976 TGAACATGAGCCGAGCAAGTATATGCTGCCACACTCTGCTGGAACA 1025  
 301 SerValGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGlu 317  
 1026 TCTGTCGAAATATTTCTACACTGGAGCCAGCAGATGTAATCTGTA 1075  
 317 uArgAlaPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsn 334  
 1076 CCGGGCAATTTGACTGGGGAGTGAACCAAAATCTGAAAAAATGCAATC 1125  
 334 InProThrProValArgTyrArgValArgAspMetThrValProThrAla 350  
 1126 AGCCACTCTCTGTAAGTACAGAGTCAAGATATAGCGTCCCTACACA 1175  
 351 MetTrpThrGlyGlnAspTrpLeuSerAsnProGluAspValLys 367  
 1176 ATGTGACAGAGAGTCAAGACTGGCTTCAATCAGAAAGACGTAAAT 1225  
 367 tLeuLeuSerGluValThrAsnLeuIleTyrHisLysAsnIleProGlu 384  
 1226 GCTGCTCTAGGTGACCAACCTCATCTACCAAGAAATATCTCTGAT 1275  
 384 rPAlaHisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyr 400  
 1276 GGGCTCAGTGGATTTCTATCTGGGGTTGGATGCTCTCACCCGATG 1325  
 401 AsnGluIleIleHisLeuMetGlnGlnGluLysThrAsnLeuSerGln 417  
 1326 AATGAATCATCATCTGATGACAGAGAGAGACCAACTTCTCCAGGG 1375  
 417 yArgCysGlnAlaValLeu 423  
 1376 ACGGTGTGAGGCCGTATTTG 1394

seq\_name: /SIDSR/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH02906  
 seq\_documentation\_block:  
 ID AAH02906 standard; DNA; 2493 BP.  
 XX  
 AC AAH02906;  
 XX  
 DT 15-JUN-2001 (first entry)  
 XX  
 DE Human shear stress-response coding sequence SEQ ID NO: 65.  
 XX  
 KW Human; shear stress-response protein; vascular disease;  
 KM arteriosclerosis; ds.  
 XX  
 OS Homo sapiens.

XX K0200125427-A1.  
XX  
XX 12-APR-2001.  
XX  
XX 02-OCT-2000; 2000MO-JP06840.  
XX  
XX 01-OCT-1999; 99JP-0280976.  
XX  
XX (KYOWA) KYOWA HAKKO KOGYO KK.  
XX (NOJI/) NOJIMA H.  
XX  
XX Nojima H., Yoshisue H., Obayashi M., Ota T., Kawabata A., Sakurada K.,  
XX Kuga T., Sekine S., Nakamura Y., Sugano S;  
XX  
XX WPI; 2001-266308/27.  
XX  
XX P-PSDB; AAB90783.  
XX  
XX  
XX DNA sequences, proteins encoded by them and antibodies against them  
XX useful in diagnosis and treatment of vascular disease caused by  
XX arteriosclerosis -  
XX  
XX  
XX Claim 20; Page 398-402; 678pp; Japanese.  
XX  
XX The present invention provides the protein and coding sequences of a  
XX number of human shear stress response proteins. These are useful in the  
XX diagnosis, treatment and screening of vascular diseases caused by  
XX arteriosclerosis, including heart failure, post-PTCA restenosis and  
XX hypertension.  
XX  
XX  
XX Sequence 2493 BP; 692 A; 484 C; 516 G; 801 T; 0 other;

## alignment\_scores:

Quality:	1306.00	Length:	410
Ratio:	3.841	Gaps:	2
Percent Similarity:	82.927	Percent Identity:	60.732

## alignment\_block:

US-09-578-063-47 x AAB02906 ..

Align seg 1/1 to: AAB02906 from: 1 to: 2493

```
2 LeuGIuTrHLeuSerArgInTrPLeuValSerHisArgMetGluMet 18
8 CTCGAGACAGCGCGCCGCGAG.....GACAGCTCCAGATGAATGCG 51
18 pLeuLeuLeuValAlaIaTyMetCPheGlnArgAsnValAsnSerValH 35
52 GTTCTGGGGTGGTGGTCTGTCTGTCTGTGGCCCTGCATCTCGAG 101
35 ISMetProThrLys.....AlaValAspProGluAlaPheMetAsnIle 49
102 GGTCTGGAGGAGAACTGACAGCTGTGATCTCTGAACAACATGTGATG 151
50 SerGluIleIleGlnHisGlnIleTyTrProCysGlnGluTyGluValAl 66
152 AGTGAATTAATCTTACTGCGGATTCCTACTAGAGATACCTAGTGA 201
66 aThrLysProGlyTyTrIleLeuSerValAsnArgIleProArgGlyLeuV 83
202 GACACAATAATGATATATCTGTGCTTAACGAAATTCCTCATGGAGAGA 251
83 aGlnProLysLysThrGlySerArgProValIleLeuGlnHisGly 99
252 AGAACCATTCACAAAGTCCCAACCAAGTGTCTCTCTGCAACATGCG 301
100 LeuValGlyGlyAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerIle 116
302 TTGCTGGCAGATTCATGTAACGTGCTCAAACTTGCCACACAGACGCT 351
116 uGlyPheIleLeuAlaAspAlaGlyPheAspAlaTrpMetGlyAsnSerA 133
```

```
352 GGCCTCATCTCTGCTGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGT 401
133 rGlyAsnAlaTrpSerArgLysHisLysThrLeuSerIleAspGlnAsp 149
402 GAGGAATACCTGCTCGTGGAAACATGAAGACACTCTCAGTTCTCAGAT 451
150 GluPheTrpAlaPheSerTyTrAspGluMetAlaArgPheAspLeuProAl 166
452 GAATCTGGGCTTTCAGTTATGATGATGAGATGGCAAAATATGACCTACAGC 501
166 aValIleAsnPheIleLeuGlnLysThrGlyGlnLysIleTyTrTyV 183
502 TTCATTAATCTCATCTGATAAATACTGGCCAAACAGAGTGTATTATG 551
183 aGlyTyTrSerGlnGlyThrMetGlyPheIleAlaPheSerThrMet 199
552 TGGTCATCTTCACAGCACACATATAGTGTATTATGCAATTTTCACAGATC 601
200 ProGluLeuAlaGlnLysIleLysMetTyTrPheAlaLeuAlaProIleAl 216
602 CCTGAGCTGGCTAAAGATTAATGTTTGGCTGGCTGGCTGGCTGGCTGGC 651
216 aThrValLysHisAlaLysSerProGlyThrLysPheLeuLeuLeuProA 233
652 TTCGTCGCCCTTCTGTAAGCTATGAGCAAAATTAAGACGATTAACAG 701
233 sPheMetIleLysGlyLeuPheGlyLysGlyGlnPheLeuTyTrGlnTr 249
702 ATCATCTCATTAAGACCTTATTGGAGACAAAGATTTCTTCCACAGAT 751
250 ArgPheLeuArgGlnLeuValIleTyTrLeuGlyGlnValIleLeuAs 266
752 GCGTTTGAAGTGGTGGGTACCCAGCTTTCACATCATGTCATATGAA 801
266 pGlnIleCysSerAsnIleMetLeuLeuGlyGlyPheAsnTrpAsnA 283
802 GGAGCTGTGGAAATCTCTGTTCTCTGTGGTGAATTAATGAGAGA 851
283 sMetAsnMetSerArgAlaSerValTyTrAlaAlaHisThrLeuAlaGly 299
852 ATTAAATATGCTAGAGTGGATGTATATACACACATTTCTCTGCTGGA 901
300 ThrSerValGlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyG 316
902 ACTCTGTGCAAAACATGTATACATGAGCCAGCTGTTAAATTCGAA 951
316 uLeuArgAlaPheAspTrpGlySerGluThrLysAsnLeuGluTyCysA 333
952 GTTCAAGCCTTGTACTGGGGAGAGAGCCAAAGATTAATTTTCATTACA 1001
333 sGlnIleProThrProValArgTyTrArgValArgAspMetThrValPro 349
1002 ACCAAGATTAATCTCCACATACATGTAAGAGATGCTTGGCCGACT 1051
350 AlaMetTrpThrGlyGlyLeuAspTrpLeuSerAsnProGluAspValIy 366
1052 CGAGCTGAGCGGGGCTGACAGCTGCTGGAGATGCTACAGAGTCAAA 1101
366 sMetLeuLeuSerGluValThrAsnLeuIleTyTrHisLysAsnIlePro 383
1102 TATCTTACTGACTGACATCACCAACTGTGTTCATGAGACATTCGCG 1151
383 IuTrpAlaHisValAspPheIleTrpGlyLeuAspAlaProHisArgMet 399
1152 AATGGAGACATCTTACATTCATTTGGGCTGTGAGATGCCCTTGGAGGCT 1201
400 TyTrAsnGluIleIleHisLeuMetGlnGln 409
1202 TATAATAAATTAATTAATTAATGAGAAA 1231
seq_name: /SIDSB/gcgdata/geneseq/geneseq/NA1993.DAT:AA042310
seq_documentation_block:
```





```

315 GlyIleuArgAlaPheAspTrpGlySerGluThrLysAsnLeuGluLys 331
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
904 GGGAAATTTCAAGCTTCACTGAGGAAAGCCCAAGACATGTTACA 953
331 scyAsnGlnProThrProValArgValArgValArgAspMetThrValP 348
954 CTACAAACCAAAAAGCCCTCCTGAATATGATGTGACAGCCATGATGTC 1003
348 rorHAlaMetTrpThrGlyGlyLysAspTrpLysSerAspProGluAsp 364
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1004 CAGTTGAGGTGGAGCGTGGCAATGACATGCGGTGATGCCCAAGAT 1053
365 ValIysMetLeuSerGluValThrAsnLeuIleuTrpHisLysAsnI 381
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1054 GTGCCATGCTGCTTCCCAACTCTCCACCTCTGTTCCATAAGGAGAT 1103
381 eProGluTrpAlaHisValAspPheIleuTrpGlyLeuAspAlaProHis 398
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1104 TCTTGCTTACATGACCTGAGCTGATGTCATGCGCAATGATGCCCTCA 1153
398 rymetTrpAsnGluIleuIleuHisLeuMetGlnGlnGluGluThrAsn 413
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1154 AGGTTTACATGAGATGATTTCCATGATGCGCAAGACTAAAGAAAT 1200

seq_name: /SID8/gcgdata/geneseq/geneseqn/NA1994.DAT:AA068388
seq_documentation_block:
ID AA068388 standard; DNA; 1528 BP.
AC AA068388;
XX
XX 20-FEB-1995 (first entry)
XX
XX Canine gastric lipase.
XX
XX Canine gastric lipase; GGL; dog; fat; mucoviscidosis; enzyme;
XX bioconversion; exocrine pancreatic insufficiency; ss.
XX
XX Canis familiaris.
XX
XX Key Location/Qualifiers
XX CDS 1..1140
XX FT /*tag= a
XX FT misc_RNA 1..1137
XX FT /*tag= b
XX FT /*note= "claim 1"
XX
XX WO9413816-A.
XX
XX 23-JUN-1994.
XX
XX 16-DEC-1993; 93WO-FR01260.
XX
XX 16-DEC-1992; 92FR-0015201.
XX
XX (LJOU ) INST RECH JOUVEINAL.
XX
XX Benicourt C, Blanchard C, Junien J;
XX
XX WPI: 1994-217890/26.
XX
XX P-PSDB; AAR56870.
XX
XX Recombinant canine gastric lipase and nucleic acid encoding it -
XX are used for improving absorption of ingested fat, treating
XX mucoviscidosis etc. and in enzymatic bio-conversions
XX
XX Claim 1; Fig 8; 52pp; French.
XX
XX The sequence given below is the sequence of figure 8, altered
XX according to the amendments described on page 2 of the appended
XX letter.
XX
XX CGL is used to improve absorption of ingested fat, in healthy and
XX sick patients (e.g. having altered levels of gastric lipase); to

```

```

CC treat conditions associated with insufficiency (or lack) of lipases,
CC esp. mucoviscidosis or exocrine pancreatic insufficiency and partic.
CC where immobilised, for bioconversions, e.g. hydrolysis or
CC transesterification (other mammalian gastric lipases, or derivs.,
CC can be used in this application).
XX
XX Sequence 1528 BP; 394 A; 386 C; 328 G; 420 T; 0 other;
SQ
Alignment_scores:
Quality: 1141.50 Length: 372
Ratio: 3.624 Gaps: 3
Percent Similarity: 84.677 Percent Identity: 56.452
alignment_block:
US-09-578-063-47 x AA068388
Align seg 1/1 to: AA068388 from: 1 to: 1528
37 ProthrLysAlaValAspProGluAlaPheMetAsnIleSerGluIleI 53
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
19 CCCACA.....AACCTGAAGTGACATGATATTAAGTCAGATGAT 59
53 eGlnHisGlnGlyTrpProGlyGluGluThrGluValAlaThrGluAsp 70
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
60 CACCTACTGGGGATACCCAGCTGAGAGATATGAGTTGTGACGAGACG 109
70 LYTyrIleLeuSerValAsnArgIleProArgGlyLeuValGlnProLys 86
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
110 GTTATATCCTTGCGATGACAGAAATTCCTTATGGAGAGAAAAATTGACAG 159
87 LysThrGlySerArgProValValLeuLeuGlnHisGlyLeuValGlyI 103
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
160 AATATAGCCCGAGACCTGTGATTTTTCACACACGCTTTGCTGCAATC 209
103 ValSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleI 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
210 AGCCACAACTGATCTTCACTGAGCTGCCACACAGCCTGCTTCATCC 259
120 euAlaAspAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAla 136
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
260 TGGCCGAGCGCGGAGACGCTGCGTGGGGAAGAGAGGAGGCAACAC 309
137 TrpSerArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAl 153
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
310 TGGGCCAGAGAGAACTGCTACTGCGCCGACCTGCGGATTCCTGGGC 359
153 aPheSerTrpAspGluMetAlaArgPheAspLeuProAlaIleAsnP 170
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
360 TTTCAAGCTTGACGAGATGCTAAATATGACCTTCCGCGCACATGTGACT 409
170 heIleLeuGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSer 186
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
410 TCATCTTGAAACAAAACGGAGACAGCAAGCTACACTACGTTGGCCATTCC 459
187 GlnGlyThrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAl 203
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
460 CAGGCGACACCACTGATTGCTTCATCGCTTCCACCAATCCCAAGCTGGC 509
203 aGlnLysIleLysMetLysPheAlaLeuAlaProIleAlaThrValLysH 220
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
510 GAAAGGATCAAAACCTTATGATTAAGCTCCGTTGCCACCTGGAAGT 559
220 IAlaLysSerProGlyThrLysPheLeuLeuLeuProAspMetMetIle 236
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
560 ACACCGAAACCCCTGTTAAACAACTCATGCTGCTGCTTCTCTCTTC 609
237 LysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArgPheLeuAr 253
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
610 AAGCTTATATTTGGAACAAAATATTC...TACCCACACACACTCTTGTGA 656
253 gGln...LeuValIleTyrLeuGlyGlnValIleLeuAspGlnIleC 269
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

657 TCAATTTCTCCGACACGAGATGCTCCCGAGACGGTGCATCTCCT 706  
 269 ySersaIlleMeleuLeuEnglylPheasnthAsnaMeLsn 285  
 707 GCAGACAGCCCTGTTATCATTTGTGATTTGACCATATGAACTTGAAC 756  
 286 MetSerATgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThSerVa 302  
 757 ATGAGTCCCTTGATGTATCTGTACATATATCCAGCATGAACTTCGGT 806  
 302 lGlnaIlleuHisTrpSerGlnAlaValaIsnSerGlyLeuEuaAga 319  
 807 TCAGACAGTGTCCACGTGTCACAGCGCTGTAAAGTCTGGAAAGTCCAG 856  
 319 lApheAsPTpGlySerGluThrLysAsnLeuGlyLysCysAsnGlnPro 335  
 857 CTTTGTAGCTGGGGAAGCCAGCTTCAGAACATGATGCATATCAGAGAC 906  
 336 ThrProValArgTyrArgValArgAspMetThrValProThrAlaMetTr 352  
 907 ATGCTCCCTACTACACCTGACAGACATGATGCGCAATCGCATGTG 956  
 352 pThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuL 369  
 957 GACGCGTGGACAGACACTGCTGGCCGACCTCAGCATGTGACCTTTGC 1006  
 369 euserGluValThrAsnLeuIleTyrHisLysAsnLeuProGluTrrpAla 385  
 1007 TTTCACAGCTCCCAATCTCATTTACACAGGAAGATCTTCCTTACAAAT 1056  
 386 HisValAspPheIleTrrpGlyLeuAspAlaProHisArgMetTyrAsnG 402  
 1057 CACTTGACCTTATCTGTGGCCATGATGCCCTCAAGCGTTTACAAATGA 1106  
 402 uIleIleHisLeuMet 407  
 1107 AATGCTTCCATGATG 1122

seq\_name: /SID58/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT58915

seq\_documentation\_block:  
 ID AAT58915 standard; cDNA to mRNA; 1528 BP.  
 AC AAT58915;  
 DT 19-AUG-1997 (first entry)  
 DE Dog gastric lipase coding sequence.  
 KW Dog; duodenal; gastric; lipase; transgenic; plant; recombinant; extract;  
 food; absorption; fat; pancreatic; cystic fibrosis; exocrine; dairy;  
 hydrolysis; trans-esterification; substrate; enzyme; biofuel; ss.  
 OS Canis familiaris.  
 FH Key  
 FT Location/Qualifiers  
 CDS 1..1140  
 /tag- a  
 /product- gastric lipase

W09633277-A2.  
 24-OCT-1996.  
 19-APR-1996; 96MO-FR00606.  
 20-APR-1995; 95FR-0004754.  
 (BIOC-) BIOCEM SA.  
 (LJOU) INST RECH JOUVEINAL.  
 Baudino S, Benicourt C, Cudrey C, Gruber V, Lenee P;  
 Merot B;

XX  
 DR WPI; 1996-485783/48.  
 DR P-PSDB; AAM09382.  
 XX  
 PT Recombinant human or canine pre-duodenal lipase prodn. in transgenic  
 PT plants - useful for facilitating absorption of fat, as bio-catalysts  
 PT and for prodn. of bio-fuel  
 XX  
 PS Claim 2; Fig 1; 130pp; French.  
 PS  
 CC This is the nucleotide sequence encoding the dog pre-duodenal (1.e.  
 CC gastric) lipase enzyme. The sequence can be used to generate transgenic  
 CC plants producing recombinant lipase in an enzymatically active form.  
 CC Alternatively the first 12 bases of this sequence (encoding amino acids  
 CC 1-4 of the protein encoded by this sequence) or bases 1-162 of this  
 CC sequence (encoding amino acids 1-54) can be deleted to form the  
 CC derivatives designated delta-4 or delta-54 respectively. Plants, or  
 CC their extracts, expressing the lipases or the truncated derivatives, can  
 CC be used: (a) as pharmaceuticals or food to facilitate absorption of fat,  
 CC either in healthy subjects or in patients with inadequate levels of  
 CC gastric/pancreatic lipase e.g. those with cystic fibrosis or exocrine  
 CC pancreatic insufficiency, in the elderly or in patients undergoing medical  
 CC treatment which alters fat absorption; (b) for performing industrial or  
 CC agricultural reactions, e.g. in processing of fats or in the dairy  
 CC industry, for hydrolysis or transesterification reactions, etc., where  
 CC the plant material may provide both enzyme and substrate. The transgenic  
 CC plants can also be used for biofuel production.  
 SO  
 Sequence 1528 BP; 392 A; 387 C; 329 G; 420 T; 0 other;

alignment\_scores:  
 Quality: 1141.50 Length: 372  
 Ratio: 3.624 Gaps: 3  
 Percent Similarity: 84.677 Percent Identity: 56.452

alignment\_block:  
 US-09-578-063-47 x AAT58915 ..

Align seg 1/1 to: AAT58915 from: 1 to: 1528

37 ProThrLysAlaValaAspProGluAlaPheMetAsnIleSerGluIleI 53  
 19 CCCACA.....AACCTGAAGTGAACAATATATATGATATGATATAT 59  
 53 eGlnHisGlnGlyTyrProCysGluGluTyrGluValaIaThrGluAspG 70  
 60 CACCTACTGGGGATRACCACTGAGAGATATGAACTTGTGACCGAAGACG 109  
 70 lYTyrlleLeuSerValaAsnArgIleProArgGlyLeuValaGlnProLys 86  
 110 GTTATATCTTGGGATCGACAGAATTCCTTATGGAGAGAAAAATTCAGAG 159  
 87 LysThrGlySerArgProValaValleuLeuGlnHisGlyLeuValaGly 103  
 160 AATATAGCGCGGAGACCTGTGCATTTTTCGAACACGGTTGCTGCAATC 209  
 103 YAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleL 120  
 210 AGCCACAACCTGATCTCCACACTGCCACACACAGCCTGCGCTTCATCC 259  
 120 eValaAspAlaGlyPheAspValaIrrpMetGlyAsnSerArgGlyAsnAla 136  
 260 TGCGCGACGCGGGATGAGCGTGTGGTGGGGAACAGCAGGGGCAACACC 309  
 137 TrpSerArgLysHisLysThrLeuSerLleAspGlnAspGluPheTrpAl 153  
 310 TGGCGCAGAGGAATCTGTACTACTCGCCGACCTCGTCAATTCGGGCG 359  
 153 aPheSerTyrAspGluMetAlaArgPheAspLeuProAlaValaIleAsnP 170  
 360 TTTCAGCTTACAGATGAGTAAATATATGACCTTCCGCCACATGATGACT 409

```

170 heileuGlnlyThrGlyGlnGluLysIleTyrValGlyTyrSer 186
187 GlnGlyThrMetGlyPheIleAlaPheSerThrMetProGluLeuAl 203
460 CAGGACACACCATGTGTTTCATCGCCTTTCCACCAATCCAGAGTGC 509
203 aglnlySileuMetTyrPheAlaLeuAlaProIleAlaThrValLys 220
510 GAACCGATCAAAACCTCTATCATAGCTCCCGTGCACCGTGAAGT 559
220 IsAlaLysSerProGlyThrLysPheLeuLeuProAspMetMetIle 236
560 ACACCGAAACCTGTTAAACAACTCATGCTCCCTTCCTGTCCTTC 609
237 LysGlyLeuPheGlyLysGlyLysGlyPheLeuTyrGlnThrArgPheLeuAl 253
610 AACGTTATATTGGAAACAAATATTC...TACCCACACACCTTTTGA 656
253 ggl...LeuValIleTyrLeuGlyGlnValIleLeuAspGlnIleC 269
657 TCATATTTCGCCACGAGGATATGTCGCCGAGAGCGGTGATCTCT 706
269 ySerAsnIleMetLeuLeuGlyGlyPheAsnThrAsnMetAsn 285
707 GCAGCAACGCCCTTATCATTTGATGATGACACTATGACCTGAAC 756
286 MetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaLysThrVal 302
757 ATGATGCGCTGGATGATGATCTGTCACATTAATCCAGACGAAACATCGGT 806
302 IGlAsnIleLeuHisTyrPheSerGlnAlaValAsnSerGlyLysLeuAla 319
807 TCAGAACGTCCTCCACTGTCGAGGCTGTTAAGTCTGGAGAGTCCAG 856
319 IapAspTyrGlySerGlyThrLysAsnLeuGluLysCysAsnGlnPro 335
857 CTTTGTACTGGGAGAGCCAGTTCAGAAACATGATGCATTCATCAAGC 906
336 ThrProValArgTyrArgValArgAspMetThrValProThrAlaMetTr 352
907 ATGCTTCCTCTACACCTGACAGACATCATGATGCCAATCGAGTGTG 956
352 pThrGlyGlyGlnAspTyrPheSerAsnProGluAspValLysMetLeu 369
957 GAGCGGTGGAGACACTGCTGCGCAGCCCTCAAGATGTTGACCTTTGC 1006
369 euSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTyrAla 385
1007 TTTCCAGAGCCCCCAATCTCATTTACACAGAGAAATCTCTTACAAAT 1056
386 HisValAspPheIleTyrGlyLeuAspAlaProHisArgMetTyrAsnG 402
1057 CACTTGACTTATCTGGCCATGATGCCCTCAACGGGTTACAAATGA 1106
402 uIleIleHisLeuMet 407
1107 AATTGTTTCCATGATG 1122
seq_name: /SID88/gcdata/geneseq/geneseqn/NA1994.DAT:AA068389
seq_documentation_block:
ID AA068389 standard; DNA; 1531 BP.
AC AA068389;
XX 20-FEB-1995 (first entry)
XX
DE Canine gastric lipase (Met-CGL).
XX
KW Canine gastric lipase; CGL; dog; fat; mucoviscidosis; enzyme;
bicoconversion; exocrine pancreatic insufficiency; ss.

```

```

XX OS Canis familiaris.
XX FH Key
XX CDS 1..1143
XX FT /*tag= a
XX PN
XX MO9413816-A.
XX 23-JUN-1994.
XX 16-DEC-1993; 93WO-FR01260.
XX 16-DEC-1992; 92FR-0015201.
XX PR
XX PA (LJOU ) INST RECH JOUVEINAL.
XX PI Bencourt C, Blanchard C, Junien J;
XX DR WPI: 1994-217890/26.
XX DR P-PSDB; AAR56871.
XX PT Recombinant canine gastric lipase and nucleic acid encoding it
XX PT are used for improving absorption of ingested fat, treating
XX PT mucoviscidosis etc. and in enzymatic bio-conversions
XX PS Claim 2; Fig 8; 52pp; French.
XX
CC The sequence given below is the sequence of figure 8, altered
CC according to the amendments described on page 2 of the appended
CC letter.
CC CGL is used to improve absorption of ingested fat, in healthy and
CC sick patients (e.g. having altered levels of gastric lipase); to
CC treat conditions associated with insufficiency (or lack) of lipases,
CC esp. mucoviscidosis or exocrine pancreatic insufficiency and partic.
CC where immobilised, for bicoconversions, e.g. hydrolysis or
CC transacetylification (other mammalian gastric lipases, or derivs.,
CC can be used in this application).
XX
SQ Sequence 1531 BP; 395 A; 386 C; 329 G; 421 T; 0 other;

alignment_scores:
Quality: 1141.50 Length: 372
Ratio: 3.624 Gaps: 3
Percent Similarity: 84.677 Percent Identity: 56.452

alignment_block:
US-09-578-063-47 x AA068389 ..
Align seg 1/1 to: AA068389 from: 1 to: 1531

37 ProthrlyslAlaValAspProGluAlaPheMetAsnIleSerGluIleI 53
22 CCCACA.....AACCTGAAGTACCATGATATGAAGTTCAGATGAT 62
53 eglnHisGlnGlyTyrProCysGlnGluTyrGluValAlaThrGluAsp 70
1 CACCTACTGGGAGATACCCAGCTGAGAAATGAAGTGTGACCGAAGACG 112
70 IyTyrIleLeuSerValAsnArgIleProArgGlyLeuValGlnProLys 86
113 GTTATATCTTGGGATGACAGAAATTCCTATGAGGAGAAATTCAGAG 162
87 LysThrGlySerArgProValIleLeuGlnHisGlyLeuValGlyG 103
163 AATATAGGCCGAGACCTGTCATTTTTCACACAGCGTTGCTCGCATC 212
103 yAlaSerAsnTyrPheSerAsnLeuProAsnAsnSerLeuGlyPheIle 120
213 AGCCACAACTGATCTCCAACTGCCCAACACAGCTGCGCTTCATCC 262
120 euAlaAspAlaGlyPheAspValTyrMetGlyLysSerArgGlyAsnAla 156

```



```

|||||.....
263 TGGCCGACCCCGGTACGACGTGTGGTGGGAACAGCGGGGCAACAC 312
137 TTPSerArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrpAl 153
|||||.....
313 TGGGCGAGGAGAAATCTGACTACTGCGCCGACCTCGGTGAATTCGGGG 362
153 aPheserTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsn 170
|||||.....
363 TTTCAGCTTTGACGAGATGGCTAAATATGACCTCCCGCACCATGTGACT 412
170 heIleuGlnLysThrGlyGlnGluLysIleTyrTyrValLysTyrSer 186
|||||.....
413 TCATCTTGAGAAACGGGACGACAGACACTACCTAGCTGGCCATTC 462
187 GlnGlyThrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAl 203
|||||.....
463 CAGGCGACACCATTTGGTTTCATCGCCTTTCCACCAATCCCAAGCTGGC 512
203 aGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaTrpValLysH 220
|||||.....
513 GAAACGGATCAAAACCTCTATGACATAGCTCCGCTGCCACCGTGAAGT 562
220 IsAlaLysSerProGlyThrLysPheLeuLeuLeuProAspMetIle 236
|||||.....
563 ACACCGAAACCTGTTAAACAACATCATGCTGCTCCCTTCCTCTTC 612
237 LysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArgPheLeuAr 253
|||||.....
613 AAGCTTATATTGGAAACAAATATTC..TACCCACACCATCTCTTCA 659
253 gGln..LeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleC 269
|||||.....
660 TCAATTTCTCGCCACCGAGGTATGCTCCCGCAGACGGAGGATCTCTCT 709
269 ysSerAsnIleMetLeuLeuGlnGlyLysPheAsnThrAsnMetAsn 285
|||||.....
710 GCAGCAACCCCTGTTTATCATTTGTGGATTGACACATGACTTGAAAC 759
286 MetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrServa 302
|||||.....
760 ATGACTCGCTTGATGTATCTGCACATATCCAGCAGAACTCCGT 809
302 IGlAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyLysLeuArga 319
|||||.....
810 TCAGAACGTGCTCCACTGCTCCAGGCTGTTAAGTCTGGGAAGTCCAA 859
319 laphasPTripGlySerGluThrLysAsnLeuGlnLysCysAsnGlnPro 335
|||||.....
860 CTTTGTACTGGGGAAGCCAGTTCAGAACATGATGCACTATCATCAGAC 909
336 ThrProValArgTyrArgValArgAspMetThrValProThrAlaMetTr 352
|||||.....
910 ATGCCTCCCTACTACAACCTGACAGACATGCATGCGCAATCGCAGTGTG 959
352 pThGlyGlyGlnAspTyrPleuSerAsnProGluAspValLysMetLeuL 369
|||||.....
960 GAACGGTGCAGACGACTGTGGCCGACCTCACGATGTTGACCTTTGTC 1009
369 euserGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAla 385
|||||.....
1010 TTTCACACTCCCAATCTCATTTTACACAGAGAGATTCCTTCAAT 1059
386 HisValAspPheIleTyrGlyLeuAspAlaProHisArgMetTyrAsnG 402
|||||.....
1060 CACTTGACCTTATCTGGCCCATGATGCCCTCAAGCGTTTACAATGA 1109
402 uIleIleHisLeuMet 407
|||||.....
1110 AATGTTTCATGATG 1125
seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1986.DAT.AAN60685

```

```

seq_documentation_block:
ID      AAN60685 standard; DNA; 1367 BP.
XX
AC      AAN60685;
XX
DT      08-JUN-1991 (first entry)
XX
DE      Sequence encoding pregastric lipase.
XX
KW      Precursor polypeptide; secretion vector; enzyme; ss.
XX
OS      Homo sapiens.
XX
FH      Key
FT      sig_peptide 47..103
FT      mat_peptide 104..1243
FT      /*tag- a
FT      /*tag- b
XX
WO8603778-A.
XX
PD      03-JUL-1986.
XX
PE      23-DEC-1985; 85WO-GB00599.
XX
PR      21-DEC-1984; 84GB-0032483.
PR      23-DEC-1985; 85WO-GB00599.
PR      01-JAN-1986; 86GB-0019568.
XX
PA      (BREM-) BREWING RES FOUND.
PA      (TUBB/) TUBB R S.
XX
PI      Tubb RS;
XX
DR      WPI: 1986-182910/28.
DR      P-PSDB; AAP60724.
XX
PT      New precursor polypeptide of defined sequence - and corresp. DNA
PT      used to transform hosts for prodn. of the polypeptide
XX
PS      Example; Fig 5; 60pp; English.
XX
CC      The inventors claim a DNA sequence encoding AAP60723 linked to a
CC      promoter upstream and a gene for a polypeptide downstream.
CC      Particular examples are the yeast emzyme AMG, the mammalian enzyme,
CC      gastric lipase and the mammalian lymphokine, interferon-alpha2.
XX
SQ      Sequence 1367 BP; 397 A; 284 C; 269 G; 417 T; 0 other;

alignment_scores:
Quality: 1124.00 Length: 405
Ratio: 3.415 Gaps: 2
Percent Similarity: 81.235 Percent Identity: 52.099

alignment_block:
us-09-578-063-47 x AAN60685 ..
Align seg 1/1 to: AAN60685 from: 1 to: 1367
14 ArgMetGluMetTyrPleuLeuIleLeuValAlaTyrMetPheGlnArgAs 30
||| ::::::::::::::: ::::| :::::
38 AGTTCCAAAATGTGGCTGCTTTTAAACATGGCAAGTTGATA..... 79
30 nValAsnSerValHisMetProThrLysAlaVal..... 41
||||| :::::
80 .....TCTGTACTGGGACTACACATGATGTTTGGAAATATACATC 122
42 .....AspProGluAlaPheMetAsnIleSerGluIleGlnHisGln 56
::::::::::::: ::::::::::::::: ::
123 CTGGAAGCCCTGAAGTGTGACTATGAACATTTATGACATATTAATTATGG 172
57 GlyTyrProCysGluGluTyrGluValAlaThrGluAspGlyTyrIleLe 73

```





PT plants - useful for facilitating absorption of fat, as bio-catalysts  
 PR and for prodn. of bio-fuel

XX Claim 3; Fig 4; 130pp; French.

XX This is the nucleotide sequence encoding the human pre-duodenal (1.e.  
 CC gastric) lipase enzyme. The sequence can be used to generate transgenic  
 CC plants producing recombinant lipase in an enzymatically active form.  
 CC Alternatively bases encoding amino acids 20-23 or 20-73 (1-4 or 1-54 of  
 CC the mature protein, respectively) can be deleted to form the derivatives,  
 CC designated delta-4 or delta-54 respectively. Plants, or their extracts,  
 CC expressing the lipases or the truncated derivatives, can be used:  
 CC (a) as pharmaceuticals or food to facilitate absorption of fat, either  
 CC in healthy subjects or in patients with inadequate levels of  
 CC gastric/pancreatic lipase e.g. those with cystic fibrosis or exocrine  
 CC pancreatic insufficiency, in the elderly or in patients undergoing medical  
 CC treatment which alters fat adsorption; (b) for performing industrial or  
 CC agricultural reactions, e.g. in processing of fats or in the dairy  
 CC industry, for hydrolysis or transesterification reactions, etc., where  
 CC the plant material may provide both enzyme and substrate. The transgenic  
 CC plants can also be used for biofuel production.

XX Sequence 1367 BP; 397 A; 284 C; 269 G; 417 T; 0 other;

# alignment\_scores:

Quality: 1124.00 Length: 405  
 Ratio: 3.416 Gaps: 2  
 Percent Similarity: 81.235 Percent Identity: 52.099

alignment\_block:  
 US-09-578-063-47 x AAT58916 ..

Align seg 1/1 to: AAT58916 from: 1 to: 1367

14 ArgMetCLeuMetThrLeuLeuLeuValAlaTyrMetPheGlnArgAs 30  
 38 AGGTCCAAATATGGCTGCTTTTACAAATGGCAAGTTGATA..... 79  
 30 nValAsnSerValHisMetProThrIlyAlaVal..... 41  
 80 .....TCTGTACTGGGACTACATGTTGTTGGAAATATACATC 122  
 42 ....AspProGluAlaPheMetAsnIleSerGluIleIleGlnHisGln 56  
 123 CTGGAAGCCCTGAGTACTATGACATGATGATGATGATGATGATGATG 172  
 57 GLeuTyrProCysGluGluIlyValAlaThrGluAspGluTyrIleLe 73  
 173 GGATACCCAAATGAGAAATATGAGTGTGCTGAGAGATGTTATATCT 222  
 73 uSerValAsnArgIleProArgGlyLeuValGlnProIlySerThrGlys 90  
 223 TGAAGTCAATATGAAATCTTATGGAGAAATAATTCAGGGAATACAGGCC 272  
 90 eArgProValIleLeuLeuGlnHisGlyLeuValGlyIlyAlaSerAsn 106  
 273 AGAGACCTGTGTGTTTTCACAGATGTTGCTTCATCAGCCCAAC 322  
 107 TrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAspAl 123  
 323 TGAATTTCCAACTGCGGAAACAGCCCTTGCTTCATCTGGGAGATGC 372  
 123 acGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerArgL 140  
 373 TGGTTATGAGTGTGGCTGGCAACAGCAGAGAAACACCGGCCAGAA 422  
 140 ySHsIlySThrLeuSerIleAspGlnAspGluPheTrpAlaPheSerTyr 156  
 423 GAAACTGTACTATTCACCAAGATTCAGTGAATCTGGGCTTCAGCTTT 472  
 157 AspGluLeuAlaArgPheAspLeuProAlaValIleAsnPheIleGlu 173

473 GATGAATGGCTAAATATGACCTTCAGCCACCAATGACTTCATGTA 522  
 173 nLysThrGlyGlnGluIlyTyrValGlyTyrSerGlnGlyThr 190  
 523 GAAACTGAGACAGAGACGACTATGTGGCAATCCAGGACCA 572  
 190 hMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLysIle 206  
 573 CCATGGTTTATGCTTTTCCACCAATCCACCGCGGCAAAAGATC 622  
 207 LysMetYrPheAlaLeuAlaProIleAlaThrValIlySHsAlaLysSe 223  
 623 AAACCTTCATGCTTACGCTCCGTTGCCACCTGTGAAGTATACAAAAG 672  
 223 rProGlyThrIlyPheLeuLeuLeuProAspMetIleLysGlyLeuP 240  
 673 CCTTATTAACAACTAGATTGTGCTCAATCCCTTCAAGTTTAT 722  
 240 heGlyLysGluPheLeuTyrGlnThrArgPheLeuArgLLeuVal 256  
 723 TTGGTGAACAAATATTTCTACCCACACACTTTTGATCAATTTCTGCT 772  
 257 IleTyrLeuCysGlyGlnValIleLeuAspGlnIleCysSerAsnIle 273  
 773 ACTGAAGTGTGCTCCGCTGAGATGCTGAATCTCTTGACAGCATGCT 822  
 273 tLeuLeuLeuGlyGlyPheAsnThrAsnMetAsnMetSerArgAla 290  
 823 ATTATATATTTGTGGATTGTGACACTGAAGACTTAAACGAGTCGCTGG 872  
 290 eValTyrAlaAlaHisThrLeuAlaGlyThrSerValGlnAsnIleLeu 306  
 873 ATGTGTATCTATCATTAATCCACAGAGAACTTGTGTTCAAAATATGTC 922  
 307 HisTrpSerGlnAlaValAsnSerGlyGluLeuArgAlaPheAspTrpG 323  
 923 CATGGACCCACAGCGCTGTAGTGTGGGAAATTCACACTTATGCTGGG 972  
 323 ySerGluThrIlyAsnLeuGluIlyCysAsnGlnProThrProValArgT 340  
 973 AAGCCGATTCAGAAATAGATGATGATGATGATGATGATGATGATG 1022  
 340 yArgValArgAspMetThrValProThrAlaMetTrpThrGlyGln 356  
 1023 ACAATGTGACAGCCATGATGATGATGATGATGATGATGATGATG 1072  
 357 AspTrpLeuSerAsnProGluAspValIlyMetLeuLeuSerGluVal 373  
 1073 GACCTGTGGCTGACCCCAAGATGTTGGCTTTGCTTCCAAACATGCC 1122  
 373 rAsnLeuIleTyrHisIlyAsnIleProGluTrpAlaHisValAspPhe 390  
 1123 CAACCTTATTTACCAAGAGATTCCTTTTACAACTGATGACCTTTA 1172  
 390 lEtrpGlyLeuAspAlaProHisArgMetYrAsnGluIleIleHisLeu 406  
 1173 TCTGGGCAATGAGAGCCCTCAAGAGATTAACAATGACATTTGTTGATG 1222  
 407 MetGlnGlnGlu 411  
 1223 ATATCAGAGATAA 1237  
 seq\_name: /SID8/gcgcdata/geneseq/geneseq/NA2001.DAT.AAH57420  
 seq\_documentation\_block:  
 ID AAH57420 standard; cDNA; 1695 BP.  
 XX AAH57420;  
 AC  
 XX  
 DT 10-SEP-2001 (first entry)  
 XX  
 DE Human stomach cell specific cDNA sequence SEQ ID NO:260.  
 XX

KW Human; tissue specific; diagnosis; brain; heart; skeletal muscle;  
 KW lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;  
 KW metabolic disease; developmental disease; cytostatic; immunomodulatory;  
 KW neuroprotective; gene therapy; cancer; immunopathology; neuropathology.  
 XX  
 OS Homo sapiens.  
 PN WO200132927-A2.  
 PD 10-MAY-2001.  
 PF 02-NOV-2000; 2000WO-US30396.  
 PR 04-NOV-1999; 99US-0163508.  
 PA (INCYTE) INCYTE GENOMICS INC.  
 PI Sornasse T, Sellhammer JJ, Watson GA;  
 DR WPI; 2001-291057/30.  
 PT New cell and tissue specific polynucleotides useful for diagnosis,  
 PT prognosis or monitoring of treatments for disorders where the gene is  
 PT associated with a cancer, immunopathology or neuropathology -  
 XX  
 PS Claim 1; Page 189-190; 327pp; English.  
 CC AAH57161 to AAH57576 represent cell and tissue specific polynucleotide  
 CC sequences (I). (I) can have cytostatic, immunomodulatory and  
 CC neuroprotective activities, and can be used in gene therapy. (I) and  
 CC proteins (II) encoded by them are used in high throughput screening  
 CC assays to select DNA molecules, RNA molecules, peptide nucleic acids,  
 CC mimetics, peptides, proteins, agonists, antagonists, antibodies or  
 CC their fragments, immunoglobulins, inhibitors, drug compounds and  
 CC pharmaceutical agents. Expression of (I) in a sample indicates the  
 CC differentiation of embryonic stem cells into a tissue selected from  
 CC brain, heart, kidney, liver, lung, skeletal muscle or pancreatic  
 CC tissues. (I) and (II) are used to produce an expression profile that  
 CC defines a metabolic or developmental process, treatment, condition,  
 CC disease or disorder. The gene profile can be used for diagnosis,  
 CC prognosis or monitoring of treatments and for investigating a  
 CC predisposition to a disorder where the gene is associated with a  
 CC cancer, immunopathology or neuropathology.  
 XX  
 SQ Sequence 1695 BP; 480 A; 356 C; 389 G; 470 T; 0 other;

alignment\_scores:  
 Quality: 1119.00 Length: 405  
 Ratio: 3.401 Gaps: 2  
 Percent Similarity: 81.235 Percent Identity: 51.852

alignment\_block:  
 US-09-578-063-47 x AAH57420 ..  
 Align seg 1/1 to: AAH57420 from: 1 to: 1695

```

14 ArgMetGluMetTrpLeuLeuValAlaTyrMetPheGlnArg 30
11 :::::::::::::::::::::
31 AGGTCCAAATGTGGCTGCTTTTACAAATGCCAGTTGATA..... 72
30 nValAsnSerValHisMetProThrLysAlaVal..... 41
73 .....TCTGTACTGGGACTACAGTGTGTTGGAAATATCATC 115
42 .....AspProGluAlaPheMetAsnLleSerGluIleGlnHisGln 56
116 CTGGAAGCCCTGAGTACTATGACATATGACATGATTAATG 165
57 GATYrProCysGlnGluTyrGluValAlaThrGluAspGlyTyrIle 73
166 GGATACCCAATGACATATGAGTTGTGACTGAGAGATGTTATATCT 215
  
```

```

73 uSerValAsnArgIleProArgIleuValAlnProLysThrGlys 90
1:::|||||
216 TGAATGCAATAGAAATCTTATGAGGAAGAAATATCAGGAAATACAGGCC 265
90 eArgProValValLeuLeuGlnHisGlyLeuValGlyValAsn 106
:::|||||
266 AGAGACCTGTGTGTTTGGACAGATGGTTTCTGCACACACCTGACCAAC 315
107 TrpLleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaSpAl 123
|||||
316 TGGATTTCACCTGCCGACACACAGCCTTGCTTCATCTGCGAGATGC 365
123 aGlyPheAspValTyrMetGlyAsnSerArgIleGlyAsnAlaTyrPseArg 140
|||||
366 TGGTATGATGTGTGGCTGGCAGACAGAGAGAAACACCTGGCGCAGAA 415
140 yHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSerTyr 156
:::|
416 GAACCTGTACTATTCACCAATTCAGTTGAATCTGGGCTTCACCTT 465
157 AspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleLeuG 173
|||||
466 GATGAATGCTAAATATGACTCTCCAGCCACATCGACTCATGTATA 515
173 nLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGlyThr 190
|||||
516 GAAGCTGCAGAGAACACACTACACTATGTGGCATTCGCCAGGCGACCA 565
190 hMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLysIle 206
|||||
566 CATGTGTTTATTTGGCTTTCCACCAATCCAGCTGGCTAAAGAAATC 615
207 LysMetTyrPheAlaLeuAlaProIleAlaTyrValIleHisAlaLys 223
||
616 AAACCTTATGCTTACCTCTGTCGCACTGCAATATACAAAG 665
223 rProGlyThrLysPheLeuLeuProAspMetIleLysGlyLeuP 240
|
666 CCTTAAACAAACATGATTTGTTCCCAACCCCTCAAGTTATAT 715
240 hGlyLysGluPheLeuTyrGlnThrArgPheLeuArgIleuVal 256
|||||
716 TTGGTGACAAATATTCTACCCACACACTCTTTGATCAATTTCTTGT 765
257 IleTyrLeuCysGlyGlnValIleLeuAspGlnIleCysSerAsnIle 273
:::|
766 ACTGAAGTGTGCTCCGCTGAGATGCGATCTCTTGACGACAGCCCT 815
273 tLeuLeuGlyGlyPheAsnThrAsnAsnMetSerArgAla 290
:::|
816 ATTATTAATTTGCTGTTGACAGTAAGAACTTAAACAGATCGCTTGG 865
290 eValTyrAlaHisThrLeuAlaGlyThrSerValGlnAsnIleu 306
:::|
866 ATGTGTATCTATCACAATATCCAGCAGAACTTGTTTAAACATGTC 915
307 HisTrpSerGlnAlaValAsnSerGlyLeuArgAlaPheAspTrpAl 323
|||||
916 CATGGACCCAGCGCTTAAGTCTGGGAATTCGAAGCTTATGACGGGG 965
323 ySerGluThrLysAsnLeuGluLysCysAsnGlnProThrProValArg 340
|||
966 AAGCCAGTTCAAGATAGATGACATATGATGATGCCACCTCCTACT 1015
340 yArgValAlaArgAspMetThrValProThrAlaMetTrpThrGlyGln 356
|||
1016 ACAATGTGACACCATGATGATGATGATGATGATGATGATGATGATG 1065
357 AspTrpLeuSerAsnProGluAspValLysMetLeuLeuSerGluVal 373
||
1066 GACCTGTGCTGACCCCAAGATGTGGCTTTTGCTTCAAACTGCC 1115
373 rAsnLeuIleTyrHisLysAsnIleProGluTrpAlaHisValAspPhe 390
  
```





XX 20-APR-2000.  
 PD 15-OCT-1999; 99WO-US24206.  
 XX 15-OCT-1998; 98US-0104436.  
 PR (GEMV ) GENETICS INST INC.  
 XX  
 PA Jacobs K, McCoy JM, Lavallee ER, Collins-Racie LA, Evans C;  
 PI Werberg D, Treacy M, Bowman MR;  
 XX  
 DR MPI: 2000-317938/27.  
 XX  
 PT Isolated polynucleotides, and encoded proteins, comprising secreted  
 PT expressed sequence tags (SESTs), useful for treating various disorders  
 PT such as autoimmune, infectious, and central nervous system disorders -  
 XX  
 PS Claim 1; Page 441; 803pp; English.

XX AAA43426 to AAA45925 represent specifically claimed secreted expressed  
 CC sequence tags (SESTs), isolated from human, mouse, chicken and rat  
 CC tissue sources. The SESTs can have a range of activities depending on  
 CC the tissues they were isolated from. The activities include:  
 CC chemotactic; proliferative; immunomodulatory; haematopoietic;  
 CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;  
 CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;  
 CC antiallergic; vulnery; antitumor; osteopathic; neuroprotective;  
 CC neurotropic; antiparkinsonian; antipsoriatic; cerebroprotective;  
 CC anticonvulsant; and antidepressant. The SESTs can be used for gene  
 CC therapy and in vaccines. The SESTs are useful as probes for the  
 CC identification and isolation of full-length cDNAs and genomic DNA  
 CC molecules which correspond to the SESTs. Proteins encoded by the SESTs  
 CC are useful in assays for determining biological activity and raising  
 CC antibodies. They may be useful for treatment of autoimmune disorders  
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions  
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,  
 CC osteoporosis, osteoarthritis, central nervous system disorders  
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation  
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's  
 CC disease), tumours, bacterial, fungal and viral infections, depression and  
 CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given  
 CC in the exemplification of the present invention.  
 XX  
 SQ Sequence 617 BP; 181 A; 125 C; 139 G; 172 T; 0 other;

# alignment\_scores:

Quality: 596.00 Length: 190  
 Ratio: 3.772 Gaps: 2  
 Percent Similarity: 83.158 Percent Identity: 60.526

## alignment\_block:

US-09-578-063-47 x AAA44349 ..

Align seg 1/1 to: AAA44349 from: 1 to: 617

16 GUMETPPLLEULEUVAIAIAATYMETPHEGINARGASNAVALAS 32  
 :::::::::::::::::::::  
 54 AAATGTGGCTGCTTTAAACAATGCAAGTTGATA..... 89  
 32 nSERVALHISMETPROTHRYLSAVAL.....A 42  
 :::::::::::::::::::::  
 90 .TCTGTACTGGGACTACACATGCTGTGTGAAATATACATCCTCGAA 138  
 42 sSPFROGUALAPHMETASNLSESGUILELLEGINHISGLNGLYTYR 58  
 :::::::::::::::::::::  
 139 GCCCTGAAGTACTATGAACTTAGTCAGATTAATTATGGGAGATAC 188  
 59 PROCYSGIUGIUTYRGUVAIAIAATHTRGUASPGIYTYRILEUSERVA 75  
 :::::::::::::::::::::  
 189 CCAATGTAGATAATGAAAGTTGTGACACGTAAGATGTTATATCTTGAAGT 238

75 IASNARGTLEPROARGGLYLEUVALGINPROLYSTYSTRGLYSERARGP 92  
 :::::::::::::::::::::  
 239 CAATGAAATTCCTTATGGAGAGAAAATTCAGGGAAATACAGGCCAGAGAC 288  
 92 roVALVALLEULEUGNHISGLYLEUVALIGLYALASERASNPTRPIL 108  
 :::::::::::::::::::::  
 289 CTGTTGTGTTTTCAGCATGTTTGTCTTCATCAGCCACAAACTGGATT 338  
 109 SERASNLEUPROASNASNSETRLEUGLYPHEILLEULEALASPALAGLYPH 125  
 :::::::::::::::::::::  
 339 TCCAAACCTGCCGAACACAGCCCTTCCTTCATTCGTGCGAGATGCTGGTTA 388  
 125 easVALTRPMETGLYASNSERARGGLYASNALATRISERARGLYSHSL 142  
 :::::::::::::::::::::  
 389 TGATGTGTGCTGGGCAACAGCAGAGAAACCTGGGCGAGAAAGAACT 438  
 142 YSTRLEUSERILEASPLINASPLUPHETRPALAPHSERTYRASPGL 158  
 :::::::::::::::::::::  
 439 TGTACTATTCCACAGATTCACTTGATTCGTGGCTTTCAGCTTGTGATGAA 488  
 159 METALARGPHEASPLEUPROALAVALLASNPHEILLEUGNLTYSTH 175  
 :::::::::::::::::::::  
 489 ATGGCTAATATATGACCTTCAGCCACAAATCGACTTCATTTGAAAGAAAC 538  
 175 RGLYGLNGLUYSILERYTYRVALIGLYTYRSEGLNGLYTHRTHMETG 192  
 :::::::::::::::::::::  
 539 TGGACAGAGAGAGCTACACTATGTTGGCCATTCACAGGACACACATTTG 588  
 192 LYPHEILLEAPHESETRH 198  
 :::::::::::::::::::::  
 589 GTTTATTGCTCTTTCACC 608





OM of: US-09-578-063-47 to: Issued\_Patents\_NA.\* out\_format : pfs  
Date: Jan 31, 2002 12:05 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

-MODEL-framer-p2n.model -DEV-xlp  
-Q-/cgn2\_1/USPTO.spool/US09578063/rn1us\_30012002\_092151\_26325/app\_query.fasta\_1.488  
-Database: Issued\_Patents\_NA.\*  
-Database sequences: 351203  
-Database length: 11338999  
-Search time (sec): 90.070000  
-SEARCH-threads=1  
-WAIT -THREADS=1

## Search information block:

Query: US-09-578-063-47  
Query length: 423  
Database: Issued\_Patents\_NA.\*  
Database sequences: 351203  
Database length: 11338999  
Search time (sec): 90.070000

## score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation	..	1137
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-08-227-108-2 +	1141.50	2461.53	1.9e-129	1137	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-09-073-674-2 +	1141.50	2461.53	1.9e-129	1137	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-08-227-108-4 +	1141.50	2461.49	1.9e-129	1140	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-09-073-674-4 +	1141.50	2461.49	1.9e-129	1140	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-08-227-108-6 +	1141.50	2461.43	1.9e-129	1146	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-09-073-674-6 +	1141.50	2461.43	1.9e-129	1146	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-08-227-108-1 +	1141.50	2461.43	1.9e-129	1146	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-09-073-674-1 +	1141.50	2461.43	1.9e-129	1146	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-08-227-108-1 +	1141.50	2457.71	3.0e-129	1528	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-09-356-952-12 +	271.50	510.09	9.2e-21	43676	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-08-751-782-2 +	155.00	326.10	1.6e-10	178	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-08-925-171-2 +	155.00	326.10	1.6e-10	178	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-08-484-105-21 +	105.00	187.71	0.0083	1676	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-08-484-106-21 +	105.00	187.71	0.0083	1676	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-08-602-359A-24 +	102.00	187.29	0.0088	1041	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-08-232-519-1 +	98.50	178.16	0.0284	1167	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-09-115-446-1 +	92.00	153.99	0.6298	2521	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-09-115-446-5 +	92.00	153.99	0.6298	2521	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-08-471-119A-1 +	92.00	116.26	79.56	46899	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-08-390-000A-3 -	91.00	153.79	0.6457	2160	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-08-742-185-102 -	91.00	116.07	82.06	40328	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-08-968-542C-1 +	89.00	136.17	6.19	6027	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-08-755-559-2 -	85.00	140.54	3.53	2180	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-09-210-474-2 -	85.00	140.54	3.53	2180	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-08-610-359A-30 +	84.50	149.29	1.15	1017	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-07-707-367-1 +	84.50	137.07	5.53	2625	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-07-186-607-1 +	84.00	138.47	4.61	2160	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-07-882-329-4 +	82.50	141.46	3.14	1329	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-08-183-213-4 +	82.50	141.46	3.14	1329	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-08-446-803-4 +	82.50	140.29	3.65	1455	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-08-661-837-4 +	82.50	140.29	3.65	1455	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-08-600-656-4 +	82.50	140.29	3.65	1455	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-09-170-670-14 +	82.50	140.29	3.65	1455	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-09-193-068-9 +	82.50	140.29	3.65	1455	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-09-193-068-13 +	82.50	140.29	3.65	1455	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-09-183-412-9 +	82.50	140.29	3.65	1455	1		
/cgn2_6/ptoddata/2/1na/5B.COMB.seq:US-09-183-412-13 +	82.50	140.29	3.65	1455	1		

seq\_name: /cgn2\_6/ptoddata/2/1na/5A.COMB.seq:US-08-227-108-2

## seq\_documentation\_block:

Sequence 2, Application US/08227108  
Patent No. 5807726  
GENERAL INFORMATION:  
APPLICANT: Blanchard, Claire  
APPLICANT: Benicourt, Claude  
APPLICANT: Junien, Jean-Louis  
TITLE OF INVENTION: Recombinant Dog Gastric Lipase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESS: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/227,108  
FILING DATE: 03-APR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fanucci, Allan A.  
REGISTRATION NUMBER: 30,256  
REFERENCE/DOCKET NUMBER: 7620-033  
TELEPHONE: 212 869-8090  
TELEFAX: 212 869-8864/9741  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1137 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1137  
US-08-227-108-2  
alignment\_scores:  
Quality: 1141.50 Length: 372  
Ratio: 3.624 Gaps: 3  
Percent Similarity: 84.677 Percent Identity: 56.452  
alignment\_block:  
US-09-578-063-47 x US-08-227-108-2  
Align seg 1/1 to: US-08-227-108-2 from: 1 to: 1137  
37 ProctHlyAlaValAspProGluAlaPheMetAsnIleSerGluIleI 53  
19 CCACCA.....AACCTGAAGTGCACCATGAATATAGTCAGATGAT 59  
53 eGlnHISgInGlyTYrProCsgInGluTYrGluValAlaThrGluAspG 70  
60 CACCTACGTGGGATACCGACGATGAGGATATGTAAGTTGTGACCGAAGAG 109  
70 lYtYrIleLeuSerValAsnArgIleProArgIleuValGlnProIys 86

```

1007 TTTCCAAGCTCCCAATCTCATTTACACAGAGATTCCTCTTACAAT 1056
386 HlsValAspHeIleTrrpGlyLeuAspAlaProHisArgMetYrAsnG1 402
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1057 CACTGGACTTATATCTGGCCATGATGCCCTCAGCGGTTTACAATGA 1106
402 uIleIleHlsLeuMet 407
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1107 AATTGTTTCATGATG 1122

seq_name: /sgn2_6/ptodata/2/lna/5B_COMB.seq:US-09-073-674-2

seq_documentation_block:
; Sequence 2, Application US/09073674
; Patent No. 5998189
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
; APPLICANT: Julien, Jean-Louis
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Warner-Lambert Company
; STREET: 2800 Plymouth Road
; CITY: Ann Arbor
; STATE: Michigan
; COUNTRY: U.S.A.
; ZIP: 48105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM-PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,674
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Crisey, Todd M.
; REGISTRATION NUMBER: 37,807
; REFERENCE/DOCKET NUMBER: 5072-D1-66-TWC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 734 622-7530
; TELEFAX: 734 622-1553
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1137 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1137
; US-09-073-674-2

alignment_scores:
Quality: 1141.50 Length: 372
Ratio: 3.624 Gaps: 3
Percent Similarity: 84.677 Percent Identity: 56.452

alignment_block:
US-09-578-063-47 x US-09-073-674-2 ..
Align seg 1/1 to: US-09-073-674-2 from: 1 to: 1137

37 ProThylsAlaValAspProGluAlaPheMetAsnIleSerGluIle1 53
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
19 CCGACA.....AACCTGAAGTACCATGATATTAAGTCAGATGAT 59
53 eGlnHlsGlnGlyTyrProCysGlnGluTyrGluValAlaThrGluAspG 70

```

```

109 60 CACCTACTGGGATACCCAGCTGAGGAATGATGAGTGTGACCGAGACG
110 70 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
111 80 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
112 90 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
113 100 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
114 110 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
115 120 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
116 130 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
117 140 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
118 150 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
119 160 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
120 170 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
121 180 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
122 190 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
123 200 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
124 210 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
125 220 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
126 230 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
127 240 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
128 250 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
129 260 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
130 270 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
131 280 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
132 290 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
133 300 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
134 310 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
135 320 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
136 330 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
137 340 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
138 350 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
139 360 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
140 370 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
141 380 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
142 390 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
143 400 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
144 410 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
145 420 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
146 430 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
147 440 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
148 450 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
149 460 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
150 470 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
151 480 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
152 490 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
153 500 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
154 510 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
155 520 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
156 530 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
157 540 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
158 550 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
159 560 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
160 570 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
161 580 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
162 590 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
163 600 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
164 610 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
165 620 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
166 630 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
167 640 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
168 650 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
169 660 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
170 670 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
171 680 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
172 690 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
173 700 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
174 710 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
175 720 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
176 730 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
177 740 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
178 750 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
179 760 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
180 770 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
181 780 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
182 790 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
183 800 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
184 810 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
185 820 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
186 830 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
187 840 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
188 850 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
189 860 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
190 870 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
191 880 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
192 890 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
193 900 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
194 910 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
195 920 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
196 930 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
197 940 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
198 950 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
199 960 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
200 970 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
201 980 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
202 990 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS
203 1000 LYTITLLEUSERVALASNAITLEPROARGLYLEUVALGILPROLYS

```

```

957 GAACGGTGGCAACGACTTGCTGGCCGACCTTCACGATGTTGACCTTTGCG 1006
369 EUSERGILVALATHRALEULEULETYRHLISLYSASNLLEPROGLUTPALA 385
1007 TTTCCAAAGCTCCCAATCATTTATTCACGAGAAAGATCTCTCTTCAAT 1056
386 HVALASAPHELLETPGLYLEUASPALAPROHISARGMETYRANGL 402
1057 CACTGGACTTTATCTGGCCATGATGCCCTTCACAGCGGTTTACATGA 1106
402 ULEULEHISLEUMET 407
1107 AATGTTCCATGATG 1122

seq_name: /cgn2_6/plodata/2/lna/5A.COMB.seq:US-08-227-108-4
seq_documentation_block:
; Sequence 4, Application US/08227108
; Patent No. 5807726
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/227,108
; FILING DATE: 03-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci, Allan A.
; REGISTRATION NUMBER: 30,256
; REFERENCE/DOCKET NUMBER: 7620-033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8664/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1140
; US-08-227-108-4

alignment_scores:
Quality: 1141.50 Length: 372
Ratio: 3.624 Gaps: 3
Percent Similarity: 84.677 Percent Identity: 56.452

alignment_block:
US-09-578-063-47 x US-08-227-108-4 ..
Align seg 1/1 to: US-08-227-108-4 from: 1 to: 1140
37 ProThrLysAlaValAlaProGluAlaPheMetAsnIleSerGluIleI 53

```

```

|||||
22 CCCACA.....AACCTGAGTACATGATATTAAGTCAGATGAT 62
53 eglhlsigllytrpocysglutrygluvalalathrcgluaspg 70
63 CACCTTACTGGGATACCACCTGAGNAATGAAGTTGTGACCGAAGAC 112
70 lytyrileuservalasnarglleproargglyleuvalglnprolys 86
113 GTRATACCTTGGGATCGACAGAAATTCCTTATGGGAGAAAATTCAGAG 162
87 LysThrGlySerArgProValValLeuLeuGlnHlsGlyLeuValGly 103
163 AATATAGGCGGAGACCTGTGATTTTCCACACGCTTGCCTGCATC 212
103 yAlaSerAsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIle 120
213 AGCCACAAACTGGATCTCCAACTGCCCAACAGAGCGCTGCTTACATC 262
120 euAlaAspAlaGlyPheAspValTTrpMetGlyAsnSerArgGlyAsnAla 136
263 TGCCCGACGCGGGGTACGACGTGTGCTGGGAAACAGCAGCGGCAACAC 312
137 TrpSerArgLysHisLysThrLeuSerIleAspGlnAspGlnuphetrpal 153
313 TGCGCCAGAGGAGAACTCTACTACTGCGCCGACTCCGTCGAATTTGCGC 362
153 apHeSerTrpAspGlnMetAlaArgPheAspLeuProAlaValIleAsn 170
363 TTTCACCTTTGACGAGATGCGTAAATATGACTTCCCGCACCATGTGACT 412
170 helLeuGlnLysThrGlyGlnGlnLysIleTyrTyrValGlyTyrSer 186
413 TCATCTTGAGAAACGCGACAGACAGACACTACACTGTCGCAATCC 462
187 GlnGlyThrMetGlyPheIleAlaPheSerThrMetProGlnLeuAl 203
463 CAGGACACCACTGTGTTTCAATCGCTTTCCACCAATCCCAAGCTGGC 512
203 agLysIleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysH 220
513 GAACGAGATCAAAACCTCTATCATGATCCGCTTCCACCGTGAAGT 562
220 IsAlaLysSerProGlyThrLysPheLeuLeuProAspMetMetIle 236
563 AACCCGAAACCTGTTAAACAACTCATGCTGCTCCCTTCCTTCCTTC 612
237 LysGlyLeuPheGlyLysLysGlnupheLeuTyrGlnThrArgPheLeu 253
613 AAGCTATATTTGAAACAAATATTC...TACCCACACACTCTTTTGA 659
253 geln..LeuValIleTyrLeuGlyGlnValIleLeuAspGlnIleC 269
660 TCMAATTTCTCGCCACCGAGATGCTCCGCGAGACGTGATCTCTCT 709
269 ySerAsnIleMetLeuLeuGlyGlyPheAsnThrAsnMetAsn 285
710 GCAGCAACGCCCTGTTTATCATTTGTGATTTGACACTTGAACCTTGA 759
286 MetSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSer 302
760 ATGAGTCGCTGATGATGTGATCTGACACTATATCCAGCAGGAACTCG 809
302 IclnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyLeuLeuArg 319
810 TCAGAACGCTGCTCACTGCTCCAGGCTGTAAAGTCTGGGAAGTCCAA 859
319 laPheAspTrpGlySerGlnupheLysAsnLeuGlnLysCysAsnGln 335
860 CTTTGACTGGGGAGGCCAGTTCAAGACATGATGACATATCATCAGAGC 909
336 ThrProValArgTyrArgValArgAspMetThrValProThrAlaMet 352

```

```

910 ATGCTTCCTCTACTACAACTGACAGACATGATGTCCTCAATCCAGTGTG 959
352 pThrGlyGlyLysTrpLeuSerAsnProGlnAspValLysMetLeu 369
960 GAACGGGGAGACAGACTGCTGGCGGACCTCAACAGATGTGACTTTGC 1009
369 euSerGlnValThrAsnLeuIleTyrHisLysAsnIleProGlnTrpAla 385
1010 TTTCGAAGCTCCCAATCTCATTTACCAACAGAAATTCCTCTTCAAT 1059
386 HisValAspPheIleTrpGlyLeuAspAlaProHisArgMetLysAsn 402
1060 CACTTGACTTATCTGGGCAATGATGCTCCCTCAAGCGGTTCATCAATGA 1109
402 uIleIleHisLeuMet 407
1110 AATTGTTTCATGATG 1125

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-073-674-4
seq_documentation_block:
: Sequence 4, Application US/09073674
: Patent No. 5998189
:
: GENERAL INFORMATION:
: APPLICANT: Blanchard, Claire
: APPLICANT: Benicourt, Claude
: APPLICANT: Junien, Jean-Louis
: TITLE OF INVENTION: Recombinant Dog Gastric Lipase
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESS: Warner-Lambert Company
: STREET: 2800 Plymouth Road
: CITY: Ann Arbor
: STATE: Michigan
: COUNTRY: U.S.A.
: ZIP: 48105
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/073,674
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Clissey, Todd M.
: REGISTRATION NUMBER: 37,807
: REFERENCE/DOCKET NUMBER: 5072-D1-66-TWC
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 734 622-7530
: TELEFAX: 734 622-1553
: TELEX:
:
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1140 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1140
: US-09-073-674-4

alignment_scores:
Quality: 1141.50 Length: 372
Ratio: 3.624 Gaps: 3
Percent Similarity: 84.677 Percent Identity: 56.452

alignment_block:
US-09-578-063-47 x US-09-073-674-4

```



us-09-578-063-47 x US-08-227-108-6  
 Align seg 1/1 to: US-08-227-108-6 from: 1 to: 1146

37 ProthlyslalValasProgluaAlaPheMetAsnIleSergIleu1 53  
 19 CCCACA.....AACCTGAAGTGACATGATATATAGTCAGATGAT 59  
 53 eglInHISglnglyTyTProCysgluGluTyGluValAlaThThGluasp 70  
 60 CACCTACTGGGATACCCAGCTGAGGAATATGAGTTGATGCCGAGAACG 109  
 70 LyTyTleuSergValAsnArgIleProArgGlyLeuValGlnProlys 86  
 110 GTTATATCCTTGGGATCGACAGAAATTCCTATGGGAGAAAAATTCAGAG 139  
 87 LysThyGlySergArgProValValLeuLeuGlnHISglYleuValGly 103  
 160 AATATAGCCGCGAGACCTGTTGCATTTTGCACAACAGGTTGCTGCAGATC 209  
 103 yAlaSerAsnTrpIleSergAsnLeuProAsnAsnSergLeuGlyPheIle 120  
 210 AGCCACAACCTGATCTCCAACTGCCCAACAGAGCTGCTGCCTTCATCC 259  
 120 euAlaSpAlaGlyPheAspValTrpMetGlyAsnSergGlyAsnAla 136  
 260 TGCCCAAGCCGGGTACGAGCTGTGCTGGGGAACAGCAGGGGCAACACC 309  
 137 TrpSergArgLysHisLysThrLeuSergIleAspGlnAspGlnPheTrpAl 153  
 310 TGCGCAGAGAGAACTGTACTACTGCGCCGACTCCGTGCAATTCGTGGCC 359  
 153 aPheSergTyRAspGluMetAlaArgPheAspLeuProAlaValAlaAsn 170  
 360 TTTCACCTTTGACGAGATGGCTAAATATGACTCCCGCCACCATTGACT 409  
 170 helLeuGlnLysThrGlyGlnGlnLysIleTyTyValGlyTySerg 186  
 410 TCATCTTGAAAGAAACGGACGAGACAGACTACACTGCTGGCCATTC 459  
 187 GlnGlyThrThrMetGlyPheIleAlaPheSergThrMetProGluLeuAl 203  
 460 CAGGGACACCATGTTTCATCGGCTTTCCACCAATCCCAAGCTGGC 509  
 203 aGlnLysIleLysMetTyRPhelAlaLeuAlaProIleAlaThrValLysH 220  
 510 GAACGAGCAAAACCTCTATGCAATTAACCTCCGTCACCGTGAAG 559  
 220 lAlAlaLysSergProGlyThrLysPheLeuLeuProAspMetMetIle 236  
 560 ACACCGAAACCTGTTAAACAACTCATCTCTCCCTTCCTTCCTTC 609  
 237 LysGlyLeuPheGlyLysLysGlnPheLeuTyRglnThrArgPheLeu 253  
 610 AAGCTTATATTGGAACAAATATTC...TACCCACACCACTTCTTTGA 656  
 253 gGln..LeuValIleTyRLeuCysGlyGlnValIleLeuAspGlnIleC 269  
 657 TCAATTTCTCGCCACCGAGATGCTCCGCGAGAGCGTGAATCTCTCT 706  
 269 ySergAsnIleMetLeuLeuGlnGlyGlyPheAsnThrAsnMetAsn 285  
 707 GAGCAACCCCTGTTTATCATTTGATGATGACATATGAACTTGAAC 756  
 286 MetSergAlaSerValTyRAlaAlaHisThrLeuAlaGlyThrSerg 302  
 757 ATGAGTCGCTTGATGTATCTGCACATATCCAGCAGAACTCGGT 806  
 302 lGlnAsnIleLeuHisTrpSergGlnAlaValAsnSergGlyLuleuArg 319  
 807 TCAGAACGCTGCTCACCTGCTCCAGGCTGTTAAAGTCTGGGAGATTCCAAG 856  
 319 lAPheAspTrpGlySergIuThrLysAsnLeuGlnLysCysAsnGlnPro 335

857 CTTTGACTGGGGAAGCCAGTTCAAGACATGATGACATATCAACAGC 906  
 336 ThrProValArgTyRArgValArgAspMetThrValProThralametr 352  
 907 ATGCTGCCCTACACAACTGACAGACATGATGATGTCATCGAGTGTG 956  
 352 pThrGlyGlnAspTrpLeuSergAsnProGluAspValLysMetLeu 369  
 957 GAACGGTGGCAAGCACTGCTGGCCGACCTCACGATGTGACTTTC 1006  
 369 eusergIuValThrasnLeuIleTyHisLysAsnIleProGluTrpAla 385  
 1007 TTTCAGAGCTCCCAATCTCATTTACCAAGAGATTCCTCTTCAAT 1056  
 386 HisValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyRAsn 402  
 1057 CACTTGACTTATCTGGGCAATGATGATGCCCTCAACGGGTTTACATGA 1106  
 402 uIleIleHisLeuMet 407  
 1107 AATGTGTTCCATGATG 1122

seq\_name: /cgn2\_6/ptodata/2/lna/5B\_COMB.seq:US-09-073-674-6

seq\_documentation\_block:

Sequence 6, Application US/09073674  
 Patent No. 5998189

GENERAL INFORMATION:

APPLICANT: Blanchard, Claire  
 APPLICANT: Benicourt, Claude  
 APPLICANT: Junien, Jean-Louis  
 TITLE OF INVENTION: Recombinant Dog Gastric Lipase  
 NUMBER OF SEQUENCES: 21  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Warner-Lambert Company  
 STREET: 2800 Plymouth Road  
 CITY: Ann Arbor  
 STATE: Michigan  
 COUNTRY: U.S.A.  
 ZIP: 48105

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/073,674

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Crissey, Todd M.  
 REGISTRATION NUMBER: 37,807  
 REFERENCE/DOCKET NUMBER: 5072-DI-66-1MC  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 734 622-7530  
 TELEFAX: 734 622-1553

TELEX:

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 1146 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA to mRNA  
 US-09-073-674-6

alignment\_scores:

Quality: 1141.50 Length: 372  
 Ratio: 3.624 Gaps: 3  
 Percent Similarity: 84.677 Percent Identity: 56.452

alignment\_block:  
US-09-578-063-47 x US-09-073-674-6

Align seg 1/1 to: US-09-073-674-6 from: 1 to: 1146

[illegible]

```

319 laPhasprlpriylserglutThrLysAsnIleuGlulYcysAsnGlnPro 335
      ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
857 CTTTTCAGCTGGGGAAACCCAGTTCGAACATGATGCACACTATCATCAGGCC 906
336 ThrProValArgTyrArgValArpAspmethrThraValProHnAlaMetTr 352
      ||| ||| : : : : ||||| ||||| ||||| |||||
907 AAGCCTCCCTACTACAACCTGACACACATGCATGTGCCAATCGCAGTGTG 956
352 pThrglyglylgnAsprTpLeuSerAsnProGluAspValLysMeteUL 369
      |:::|||||::||| |||||::|||::|||::|||::|||::|||
957 GAAGGTGGCAACGACCTGTGCTGGCGGACCCTCAGCATGTGTGACCTTTGGC 1006
369 euSereGUalThrAsnIleuIeTYrHisLYsAsnIlleProGluITrpAla 385
      |||||::|||::|||::|||::|||::|||::|||::|||::|||
1007 TTTCCAGACTCCCACATCTCATTTCCACAGAAAGATTCTCTCTTACAT 1056
386 HisValAspPheIleTrpGLyLeuAspAlaProHisArGmetTyrrAsnGI 402
      |||::|||::|||::|||::|||::|||::|||::|||::|||
1057 CACTTGAGACTTATATGTGGCCATGTGATGCCCTCCCAAGCGGTTCACAArGA 1106
402 uileIleHIsIeuMet 407
      |||::|||::|||::|||::|||::|||::|||::|||
1107 AATGTGTTTCANtGAtG 1122
```

seq\_name: /cgn2\_6/ptodata/2/lna/5A\_COMB.seq:US-08-227-108-1

seq\_documentation\_block:

; Sequence 1, Application US/08227108

; Patent No. 5807726

GENERAL INFORMATION:

APPLICANT: Benicourt, Claude

APPLICANT: Junien, Jean-Louis

TITLE OF INVENTION: Recombinant Dog Gastric Lipase

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds  
ADDRESS: 1155 Avenue of the Americas  
CITY: New York, New York

STREET: 1155 AV  
CITY. New York

CITY: New York  
STATE: New York

STATE: NEW  
COUNTRY: U

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

```

;
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0 Version #1.25

```

```

; SOFTWARE: PatentIn Release #1.0,
: CURRENT APPLICATION DATA:

```

```

CORRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/227,108

```

FILING DATE: 03-APR-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Fanucci, Allan A.

REGISTRATION NUMBER: 30,256  
REFERENCE/DOCKET NUMBER: 7630-033

REFERENCE/DOCKET NUMBER: 76

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090

TELEPHONE: 212 750-3030  
TELEFAX: 212 869-8864/9741

TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1528 base pairs

TYPE: nucleic acid  
STRANDEDNESS: double

```

; STRANDEDNESS: double
; TOPOLOGY: linear

```

MOLECULE TYPE: CDNA to mRNA

US-08-227-108-1

```
alignment_scores:
```

Quality: 1141.50

Ratio:	3.624	Gaps:	3
Percent closed:	0.677	Percent traded:	55.453



alignment\_block:  
US-09-578-063-47 x US-08-227-108-1 ..

Align seg 1/1 to: US-08-227-108-1 from: 1 to: 1528

```

37 Prothrlyslalavalaspptroglualaphemetaleserjullel 53
   |||||
   |||||
19 CCCACA.....ACCCGGAAGTACCATGATATAGTCACATGAT 59
   |||||
   |||||
53 eglhislnglytyrprocgslugllyglulvalalathrgluasp 70
   |||||
   |||||
60 CACCTACGGGGATACCCAGCTGAGGAATGAACTGTGACCGAAGAG 109
   |||||
   |||||
70 lytyrlyleuservalasnargileproarglyleuvalglnpro 86
   |||||
   |||||
110 GTTATATCTCTGGGATCGACAGAAATCTTATGAGAGAAAAATTCAG 159
   |||||
   |||||
87 LysThrlySerArgProValValLeuLeuGlnHsGlyLeuValGly 103
   |||||
   |||||
160 AATATAGCCCGAGACCTGTGGCATTTTGCACACGGTTTCTGCATC 209
   |||||
   |||||
103 yAlaserntprlleserasnleuproasnserleuglypheile 120
   |||||
   |||||
210 AGCCCAAACTGGATCTCCAACTGCCCAACACAGCTGGCCCTCATCC 259
   |||||
   |||||
120 euAlasAlaGlyPheaspValTpmMetGlyAsnSerArgGlyAsn 136
   |||||
   |||||
260 TGGCCGAGCCGGGTACGACGCTGTGGGAGACAGCGGGCAACACC 309
   |||||
   |||||
137 TrpSerArgLysHsLysThrLeuSerIleaspGlnAspGlyPhe 153
   |||||
   |||||
310 TGGCCGAGAGAAATCTGATACCTGCGCCGACCTCGTGATTTCTGG 359
   |||||
   |||||
133 aPheSerTyraSpGluMetAlaArgPheaspLeuProAlaValI 170
   |||||
   |||||
360 TTTCAGCTTTGACGAGATGGCTAAATATGACCTCCCGCACCATG 409
   |||||
   |||||
170 heileleGlnLysThrGlyGlnGlyLysIleTyTrValGlyTyr 186
   |||||
   |||||
410 TCATTTTGAAGAAACGGGACAGCAACGCTACCTAGCTTGCCATTC 459
   |||||
   |||||
187 GlnGlyThrThrMetGlyPheIleAlaPheSerThrMetProGlu 203
   |||||
   |||||
460 CAGGACACCATGATGTTTATCCCTTTTCCACCAATCCCAACCTGC 509
   |||||
   |||||
203 aGlnLysIleLysMetTyrrPheAlaLeuAlaProIleAlaThr 220
   |||||
   |||||
510 GAAACGGATCAAAACCTTCTATGCTAGCTCCCTGCGACCGTGA 559
   |||||
   |||||
220 lAlaLysSerProGlyThrLysPheLeuLeuProaspMetIle 236
   |||||
   |||||
560 ACACCGAACCCTGTTAAACAACTCATGCTGCTCCCTGCTCTTC 609
   |||||
   |||||
237 LysGlyLeuPheGlyLysGlyLysGlyPheLeuTyrrGlnThr 253
   |||||
   |||||
610 AAGCTATATTGGAAACAAATATTC..TACCCACACACTTCTTTGA 656
   |||||
   |||||
253 gGln..LeuValIleTyrrLeuGlyGlnValIleLeuaspGln 269
   |||||
   |||||
657 TCAATTTCTCGCCACGAGGATGCTCCCGCAGACGGTGCATCTCT 706
   |||||
   |||||
269 ySserAsnIleMetLeuLeuGlyGlyPheAsnThrAsnMetAsn 285
   |||||
   |||||
707 GCAGCAAGCCCTGTTTATCATTTGTGGATTGACACTATGACCT 756
   |||||
   |||||
286 MetSerArgAlaSerValTyrrAlaAlaHsThrLeuAlaGlyTh 302
   |||||
   |||||
757 ATGATCGCTTGATGTGATCTGCACATATCCAGCAGCAACATCG 806
   |||||
   |||||
302 lGlnAsnIleLeuHsIleTrpSerGlnAlaValAsnSerGlyL 319
   |||||
   |||||
807 TCAGAACGTCTCACATGCTGCCAGGCTTTAAGTCTGGGAATTC 856
   |||||
   |||||

```

```

319 lApeaspTrpGlySerGlyThrLysAsnLeuGlyLysCysAsnGlnPro 335
   |||||
   |||||
857 CTTTGTAGCTGGGAGCCAGTTCAGACATGATGACATCATCATCAGAC 906
   |||||
   |||||
336 ThrProValArgTyrrValArgaspMetThrValProThrAlaMet 352
   |||||
   |||||
907 ATGCTCCCTCACTACACCTGACAGACATGATGTGCCAATGCGAGT 956
   |||||
   |||||
352 PthGlyGlyGlnAspTrpLeuSerAsnProGlnAspValLysMetLeu 369
   |||||
   |||||
957 GAGCGGTGCACAGACTTGTGCGCCGACCTGCACATGTTGACCTTTC 1006
   |||||
   |||||
369 euSerGlyValThrAsnLeuIleTyrrHsLysAsnIleProGlyTr 385
   |||||
   |||||
1007 TTTCACACTGCCCAATCTCATTTACACAGAGAAATCTCTTACAT 1056
   |||||
   |||||
386 HsValAspPheIleTrpGlyLeuAspAlaProHsAspMetTyrr 402
   |||||
   |||||
1057 CACTTGACATTATCTGGGCGCATGATGCTCCCTCAAGCGGTTTACA 1106
   |||||
   |||||
402 uIleIleHsLeuMet 407
   |||||
   |||||
1107 AATGTTTCCATGATG 1122
   |||||
   |||||
seq_name: /cgn2_6/ptodata/2/1na/5b_COMB.seq:US-09-073-674-1
seq_documentation_block:
; Sequence 1, Application US/09073674
; Patent No. 5998189
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claire
; APPLICANT: Junten, Jean-Louis
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Warner-Lambert Company
; STREET: 2800 Plymouth Road
; CITY: Ann Arbor
; STATE: Michigan
; COUNTRY: U.S.A.
; ZIP: 48105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0., Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,674
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Crissey, Todd M.
; REGISTRATION NUMBER: 37,807
; REFERENCE/DOCKET NUMBER: 5072-D1-66-TMC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 734 622-7530
; TELEFAX: 734 622-1553
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1528 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; US-09-073-674-1

```

alignment\_scores: Quality: 1141.50 Length: 372  
Ratio: 3.624 Gaps: 3

Percent Similarity: 84.677 Percent Identity: 56.452

alignment\_block:

US-09-578-063-47 x US-09-073-674-1 ..

Align seg 1/1 to: US-09-073-674-1 from: 1 to: 1528

```

37 ProthlyAlaValasProgluAlaPheMetAsnIleSerGluIle1 53
   ||||| ..... |||||.....
19 CCCACA.....AACCTGAAGTACCATGAATATGAATGATGAT 59
   |||||.....
53 eGlnHsGlnGlyTyrProCysGluGluTyrGluValAlaThrGluAsp 70
   | : : : |||||.....
60 CACCTACTGGGGATACCACTGAGGAAATATGAGTTGTCACCGAAGAC 109
   |||||.....
70 LysThrIleLeuSerValaAsnArgIleProArgGlyLeuValGlnPro 86
   |||||.....
110 GTTATATCTCTGGATCGACAGAAATCTTATGAGGAGAAATTCAGAG 159
   |||||.....
87 LysThrGlySerArgProValaIleLeuGlnHsGlyLeuValGlyI 103
   : : : |||||.....
160 AATATAGCGCGAGACCTGTGCTTTTGCACACGCTTTCCTGCATC 209
   |||||.....
103 ValaSerAsnTrpIleSerAsnLeuProAsnSerLeuGlyPheIle 120
   |||||.....
210 ACCCAAACTGATCTCCACCTGCCCAACACAGCTGCTTATCC 259
   |||||.....
120 eualaAspAlaGlyPheAspValaTrpMetGlyValaSerArgGlyAsn 136
   |||||.....
260 TGGCGGACCGCGGATCGTGTGCTGGGGAACAGAGGGCAACACC 309
   |||||.....
137 TrpSerArgLysHisLysThrLeuSerIleAspGlnAspGluPheTrp 153
   |||||.....
310 TGGGCGAGAGAAATCTGACTACTGCCCGGCTCGGAAATTCGCGGC 359
   |||||.....
153 APheserTyrAspGluMetAlaArgPheAspLeuProAlaValIleAsn 170
   |||||.....
360 TTTCACCTTGACGAGATGCTTAATATACCTCCCGCACCATGACT 409
   |||||.....
170 heIleuGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSer 186
   |||||.....
410 TCATCTTAAGAAAGAGGAGGACAGACACTACGTTGCCATTC 459
   |||||.....
187 GlnGlyThrThrMetGlyPheIleIleAspSerThrMetProGluLeu 203
   |||||.....
460 CAGGGGACACCATGCTTCACTGCTTCCACCAATCCCAAGCTGGC 509
   |||||.....
203 aGlnLysIleLysMetTyrPheAlaLeuAlaProIleAlaThrValLys 220
   |||||.....
510 GAAAGGATCAAAACCTTATGATACCTCCCGTTGCCACCGTGAAGT 559
   |||||.....
220 lAlaLysSerProGlyThrLysPheLeuLeuProAspMetMetIle 236
   |||||.....
560 AACCGAAACCTGTTAAACAACTCATGCTGCTCCCTTCCTTC 609
   |||||.....
237 LysGlyLeuPheGlyLysLysGluPheLeuTyrGlnThrArgPheLeu 253
   |||||.....
610 AAGCTATATTTGAAACAAATATATTC...TACCCACACACTTCTTGA 656
   |||||.....
253 gln...leuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIle 269
   |||||.....
657 TCATATTTCTGCCACCGAGGTATGCTCCCGGAGACGCTGATCTCTCT 706
   |||||.....
269 yASerAsnIleMetLeuLeuLeuGlyLysPheAsnThrAsnAsnMetAsn 285
   |||||.....
707 GAGAGAACCCCTGTTATCATTTGTGATTTGACACTATGAACCTGAAC 756
   |||||.....
286 MetSerArgAlaSerValaTyrAlaAlaHisThrLeuAlaGlyThrSera 302
   |||||.....
757 ATGAGTCGCTTGATGTATCTGTACATAATCCAGCAGAACATCGGT 806
   |||||.....
302 lGlnAsnIleLeuHisTrpSerGlnAlaValaAsnSerGlyLeuLeuArg 319
   |||||.....

```

```

807 TCAGAACGTCTCCACTGGTCCAGGCTGTATAGTCTGGGAAGTTCAG 856
319 lAphaSPTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnPro 335
   |||||.....
857 CTTTGTACTGGGGAGACCCAGTTACAGACATGATGCATATCATCAGAG 906
   |||||.....
336 ThrProValaArgTyrArgValaArgAspMetThrValProThrIleMet 352
   |||||.....
907 ATGCTCTCCCTACTATCAACCTGCAGACATGATGATGCCAATGCCAGT 956
   |||||.....
352 pThrGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeu 369
   |||||.....
957 GAGCGGTGGGAAGACTGTGTGGCCAGCTTACAGATGTGACCTTTGCG 1006
   |||||.....
369 euSerGluValaThrAsnLeuIleTyrHisLysAsnIleProGluTrpAla 385
   |||||.....
1007 TTTCCAAGCTCCCAATCTATTTACACAGAAAGATCTCTCTCAAT 1056
   |||||.....
386 HisValaSPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsn 402
   |||||.....
1057 CACTTGACTTATTCGGGCATGATGCCCTCAAGCGGTTTACATGA 1106
   |||||.....
402 uIleIleHisLeuMet 407
   |||||.....
1107 AATGTTTTCATGATG 1122
   |||||.....
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-356-952-12
seq_documentation_block:
; Sequence 12, Application US/09356952
; Patent No. 6117663
; GENERAL INFORMATION:
; APPLICANT: Boriack-Sjodin, Ann
; APPLICANT: Margalit, S. M.
; APPLICANT: Bor-Sogil, Dafna
; APPLICANT: Cole, Philip
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
; FILE REFERENCE: 600-1-228R
; CURRENT APPLICATION NUMBER: US/09/356,952
; EARLIER FILING DATE: 1999-07-19
; EARLIER APPLICATION NUMBER: 60/093,631
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 43676
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-356-952-12
alignment_scores:
Quality: 271.50 Length: 376
Ratio: 1.263 Gaps: 16
Percent Similarity: 57.181 Percent Identity: 25.532
alignment_block:
US-09-578-063-47 x US-09-356-952-12 ..
Align seg 1/1 to: US-09-356-952-12 from: 1 to: 43676
61 gLUGluTyrGluValaIleThrGluAspGlyTyrIleLeuSerVal...As 76
   |||||.....
14769 GAGGAATTAGATTGAGACCGAGATGGCTTTGTTATGATTTGCGCA 14818
   |||||.....
76 nArgIleProArgGlyLeuValGlnProLysLysThrGlySerArgPro 93
   |||||.....
14819 CTTGATACCAAAATATATAGACAGATCTGTGACAGAAAGAGGCCAC 14868
   |||||.....
93 alValleuLeuGlnHsGlyLeuValGlyGlyValaSerAsnTrpIleSer 109
   |||||.....

```

```

381 .....lileproglutripalahlstvalasphaeilletrpcilyueuaspl 395
      ||| ||||:|||||::||| ::|||:: |||:
15738 TGGTACATTGAGTAGATGCCCATATGTGATGCCTATGGCAGCATGATGT 15787
395 aprohlslArgmetyrAsngluille 404
      : ::|||::: :: |||::
15788 CATAGAGCAATTTGTAACCATTTTTA 15815

seq_name: /cgn2_6/plodata/2/lna/5A_COMB.seq:US-08-751-782-2

seq_documentation_block:
  Sequence 2, Application US/08751782
  Patent No 5821352
  GENERAL INFORMATION:
    APPLICANT: Heintz, Nathaniel
    APPLICANT: Gubday, Jonathan
    APPLICANT: Skinner, Michael
    TITLE OF INVENTION: A CDNA library prepared during
    TITLE OF INVENTION: Regression of Rat Prostate and Methods of Use Thereof
    NUMBER OF SEQUENCES: 15
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: David A. Jackson, Esq.
      STREET: 411 Hackensack Ave, Continental Plaza, 4th
      STREET: Floor
      CITY: Hackensack
      STATE: New Jersey
      COUNTRY: USA
      ZIP: 07601
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patentin Release #1.0, Version #1.30
      CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/08/751,782
        FILING DATE: 18-NOV-1995
      CLASSIFICATION: 435
      ATTORNEY/AGENT INFORMATION:
        NAME: Jackson Esq., David A.
        REGISTRATION NUMBER: 26,742
        REFERENCE/DOCKET NUMBER: 600-1-190
        TELECOMMUNICATION INFORMATION:
          TELEPHONE: 201-487-5800
          TELEFAX: 201-343-1684
        INFORMATION FOR SEQ ID NO: 2:
          SEQUENCE CHARACTERISTICS:
            LENGTH: 178 base pairs
            TYPE: nucleic acid
            STRANDEDNESS: double
            TOPOLOGY: linear
          MOLECULE TYPE: cDNA
          HYPOTHETICAL: NO
          IMMEDIATE SOURCE:
            CLONE: 10.2
          US-08-751-782-2

alignment_scores:
  Quality: 155.00          Length: 50
  Ratio: 3.780             Gaps: 0
  Percent Similarity: 82.000   Percent Identity: 56.000

alignment_block:
  US-09-578-063-47 x US-08-751-782-2 ...

Align seg 1/1 to: US-08-751-782-2 from: 1 to: 178

139 ATGTLGSHLSystrleuserrlleaspjnaspgluiphetralAlaphe 155
      ||||||| ||||: ||||: ||||: ||||: ||||:
       AGGAACAATGTCGCCCTAGACCACAGATTCTTAAGAATTTTGCGATT 56
155 rTyraSpjUmetAlairgPheaspleuProAlavalIleasnphelld 172
```



```

101 1GlyGlyAlaSerAsnTrpIleSerAsnLeuProAsnAsn..... 114
117 GAATGGAAGAAAGAAATGCTTTCAGAAATTTGCATCAATTTAGAGAAG 166
115 .....SerLeuGlyPhe.....IleuAlaAspAla 123
167 ATTTGGAACAACTGGCTTCAGAGATGAACCTGTATCAATGGCTCAATCA 216
134 GlyPheAspValTrp.....MetGlyAsnSer.....ArgG1 134
217 GCAATCGAAATTAATTAATTCAGAGAAATTCGGGTCAGACGAATGA 266
134 yAsnAlaTrpSerArgGlyShsLys.....ThrLeuS 145
267 TAATCGGAATCCCTCGTGGAGAGCTGCGAATGGAATGAACTGAAG 316
145 eTlleAspGlnAspGluPheTrpAlaPheSerTyrAspGluMetAlaArg 161
317 AATTTGAGAGAGACGATGACATCAATATGCTACTGATTTCAACAAA 366
162 PheAspLeuProAlaValIleAsnPheIleLeuGlnLysThrGlyGlnG1 178
367 TGTGATCTCCCTGACCTCGAAATATATATTACCAAAAAAGATACACGA 416
178 uLysIleTyrValGlyTyrSerGlnGlyThrMetGlyPheIleA 195
417 ATTCGAAACGATTCGAGCATCTCGCGATATATTCGGA..... 459
195 laPheSerThrMetProGluLeuAlaGlnLysIleLysMetTyrPheAla 211
460 .....AATGGAAGCTTATTCATCAG 480
212 leuAlaProIleAlaThrValLysHisAlaLysSerProGlyThrLysP 228
481 GCTGGA..... 486
228 eleuLeuLeuProAspMetMetIleLysGlyLeuPheGlyLysGluP 245
487 .....TTTAATATCTCTTTCACGCTGCGTTCGAGCGTGAGG 526
245 heLeuTyrGlnThrArgPheLeuArgGlnLeuValIleTyrLeuGlyS 261
527 TTCTC.....ACAGATTTGGAATGAGCTATCCGATTATACATATAG 570
262 GlnValIleLeuAspGlnIleCysSerAsnIleMetLeuLeuGlyG1 278
571 AGAGTGGATGCACGGAAGATGGCTCAATGTAAAGTTCTTCCTGGAGC 620
278 yPheAsnThrAsn.....AsnMetAsnMetSerArgAlaSerValT 292
621 TATCATATGAGATATGAGCTGAAATGTATGTGTGAAGAGCG..... 663
292 yTAlaAlaHisThrLeuAlaGlyThrSerValGlnAsnIleLeuHisTrp 308
664 .....CAATCTACGATTGTGG 681
309 SerGlnAlaVal.....AsnSerGlyGluLeu..... 317
682 GCTCATCTATTCGCAAGAAATGATAGCAACAGTTGATCTTATCAT 731
318 .....ArgAlaPheAspTrpGlySerGlu 325
732 TGATATATATGAGCTCCTGATTTGAGAGAGTGAT 765

```

seq\_name: /cgn2\_6/prodata/2/lna/5A\_COMB.seq:US-08-484-106-21

seq\_documentation\_block:

; Sequence 21, Application US/08484106

; Patent No. 5614618

; GENERAL INFORMATION:

; APPLICANT: STILLMAN, Bruce

; APPLICANT: BEILL, Stephen P

```

APPLICANT: KOBAYASHI, Ryuji
APPLICANT: RINE, Jasper
APPLICANT: FOSS, Margit
APPLICANT: McNALLY, Francis J
APPLICANT: LAURENSON, Patricia
APPLICANT: HERSKOWITZ, Ira
APPLICANT: LI, Joachim J
APPLICANT: GAVIN, Kimberly
TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOEBACH, TEST, ALABRITON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,106
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard Aron
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8770
TELEFAX: (415) 494-8771
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 1676 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 13..1302
US-08-484-106-21

```

#### alignment\_scores:

```

Quality: 105.00 Length: 278
Ratio: 0.802 Gaps: 13
Percent Similarity: 47.122 Percent Identity: 21.942

```

#### alignment\_block:

US-09-578-063-47 x US-08-484-106-21 ..

Align seg 1/1 to: US-08-484-106-21 from: 1 to: 1676

```

85 ProLysLysThrGlySerArgProValValLeuLeuGlnHisGlyLeuVal 101
||||| |||||
85 CCAGAAAAAGAGAGATCCAGA.....CAGAAAAAGAC 116
101 1GlyGlyAlaSerAsnTrpIleSerAsnLeuProAsnAsn..... 114
117 GAATGGAAGAAAGAAATGCTTTCAGAAATTTGCATCAATTTAGAGAAG 166
115 .....SerLeuGlyPhe.....IleuAlaAspAla 123
167 ATTTGGAACAACTGGCTTCAGAGATGAACCTGTATCAATGGCTCAATCA 216
124 GlyPheAspValTrp.....MetGlyAsnSer.....ArgG1 134
217 GCAATCGAAATTAATTAATTCAGAGAAATTCGGGTCAGACGAATGA 266

```

```
COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 6.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/602,359A
FILING DATE: February 16, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: HALL, LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5070
TELEFAX: 619-678-5099
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 1041 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: GENOMIC DNA
US-08-602-359A-24

Alignment scores:
Quality: 102.00 Length: 180
Ratio: 1.052 Gaps: 6
Percent Similarity: 53.889 Percent Identity: 22.778

Alignment block:
US-09-578-063-47 x US-08-602-359A-24 ..

Align seg 1/1 to: US-08-602-359A-24 from: 1 to: 1041

20 LeuileLeuValAlaIATyMetPheGlnArgAsnValAsnSerValHisMe 36
|||||:|||||:||||| ||| ||||| :|
178 CTGCCTCCGTGGCGCTGGTGTCTGTGGCGCTGTTGTCTCCTGTTGTT 227
36 tprothr.....LysAlaValAspProGluAlaPheMetAsnI 49
|||||:|||||:|||||:|||||:|||||:
228 CCGGGCTATGTGCGCATTAAGTCATGCAACCGGAGAAGTGTCGCCCG 277
49 IeserGluIleIleGlnHisIeGlnIyTyProCysGluGluTyrgLyuAl 65
|||||:|||||:|||||:|||||:|||||:
278 TTGCCCGC.....GAGGCGCTCACCTTACAAGACTTCAGCGTG 315
66 AlathrGluAspGlyTyrrIleLeuSerValAsnArgIleProArgGlyLe 82
|||||:|||||:|||||:|||||:|||||:
316 ACCGGGAGAGATGAGCTGTGTGTT.....CGGGCTG 347
82 uValGlnProLysThrIySeratgProValValLeuLeuGlnHisG 99
|||||:|||||:|||||:|||||:|||||:
348 GGTGGTGGCCCCGCGCGCTGGGGCAACCGGCTTCGTTTAAATGACG 357
99 IyleuValGly...GlyAlaSerAsnTrpIleSerAsnLeuProAsn 114
|||||:|||||:|||||:|||||:|||||:
398 GGTAACAGGGGTGGCGCGCGCGCCCAACATGCGTGTGCGCGCCGGAG 447
115 SerLeuGlyPheIleLeuAlaAspAlaGlyPheAspValTrpMetGlyAs 131
|||||:|||||:|||||:|||||:|||||:
448 .....CTTCGTGAATGGGGGAGACC CGGTGTGTGTGTTCCA 482
131 nSerArgLyAsnAlaTrpSerArgLyShisLyThrLeuSerIleAspG 148
|||||:|||||:|||||:|||||:|||||:
483 CTTCGGGGGCGACGGGAGACCGGGGCGCTCGACACGATTGG..... 525
```

148 InAspGluHepTrpAlaPheSerTyrAspGluMetAlaArgPheAspLeu 164  
 526 ..... CCGCGGAGAGTGGTGGATGCC 546  
 165 ProAlaValAlaIleAsnPheIleLeuGlnLysThrGlyGlnLysIleTyr 181  
 547 CCGGCGCTGTGGGTATGTCTCGAGAGCGGTCCCGCGCGCGATMAAT 596  
 181 TTYrValGlyTyrSerGlnGlyThrThrMetGlyPheIle 194  
 597 ATTGGTGGGTTACGTATGGCGCGCTGTACGATCGTG 636

seq\_name: /cgn2\_6/ptodata/2/1na/5A\_COMB.seq:US-08-232-519-1

seq\_documentation\_block:

Sequence 1, Application US/08232519  
 Patent No. 5484725  
 GENERAL INFORMATION:  
 APPLICANT: Kageyama, Bunji  
 APPLICANT: Nakae, Masanori  
 APPLICANT: Yagi, Shigeo  
 TITLE OF INVENTION: NORBORANE TYPE ESTER HYDROLASE  
 NUMBER OF SEQUENCES: 2  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fish & Neave  
 STREET: 1251 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10020  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/232,519  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 5-96286  
 FILING DATE: 22-APR-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Haley Jr, James F.  
 REGISTRATION NUMBER: 27,794  
 REFERENCE/DOCKET NUMBER: SHGN-4  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 596-9000  
 TELEFAX: (212) 596-9090  
 TELEX: 14-8367  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1167 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 ORIGINAL SOURCE:  
 ORGANISM: Acetobacter pasteurianus  
 STRAIN: ATCC 12873  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..1164  
 US-08-232-519-1

alignment\_scores:

Quality: 98.50 Length: 400  
 Ratio: 0.518 Gaps: 20  
 Percent Similarity: 47.500 Percent Identity: 20.250

alignment\_block:

US-09-578-063-47 x US-08-232-519-1

Align seg 1/1 to: US-08-232-519-1 from: 1 to: 1167

38 ThrLysAlaValAlaAspProGluAlaPheMetAsnIleSerGluIleLeu 54  
 250 ACCGCCAGATGTGATATGAGCCTTATC ..... 279  
 54 nHsGlnGlyTyrProCysGluGluTyrGlu ..... ValAlaThrGlu 69  
 280 ..... AAAAATTTCATGCAGAAAACCCAGACAGATGCGGTGCAGACGAAG 325  
 69 spGlyTyrIleLeuSerValAsnArgIleProArgGly ..... LeuVal 83  
 326 ATGCCTTCGCGGGGAGAACCCAGACAGGTACGTTAGCGACACACAGCTA 375  
 84 GlnProLysLysThrGlySerArg ..... ProValIleLeuLeu 97  
 376 AATGTCCTGTGATGTTCGACGACGAGGCGACGCTATGTCTGTG.. 423  
 97 nHsGlyLeuValGlyAlaSerAspTrpIleSerAsnLeuProAsn 114  
 424 .CAGGTTTGGCGGAGATATCAGCACTGGCTG ..... CTCACACAG 466  
 114 snSerLeuGlyPheIleLeuAlaAspAlaGlyPheAspValTrpMetGly 130  
 467 ATGCCTTGGCCCGACAAAGCCCGTAATTCCTTGTATCTG ..... 507  
 131 AsnSerArgGlyAsnAlaTrpSerArgLysHis ..... LysThrLe 144  
 508 ..... CCGGGCATGGGCGCTTCCTCTAAACGTCGACAGCAGCGCT 551  
 144 uSerIleAspGlnAspGluPheTrpAlaPheSerTyrAspGluMetAla 161  
 552 GCGGTTT ..... 558  
 161 rgPheAspLeuProAlaValAlaAsnPheIleLeuGlnLysThrGlyGln 177  
 559 ..... TTGCGCGGTGTGTGTAAGGAAATGCTGCACAAACCTTAATA 600  
 178 GluLysIleTyrTyrValGlyTyrSerGlnGlyThrThrMetGlyPheI 194  
 601 GAAAAAGCCCATGTGTGGCCATCTTGGGG ..... GCGGCGAT 641  
 194 eAlaPheSerThrMetProGluLeuAlaGlnLysIleLysMetTyrPhe 211  
 642 TCCCTGACCTGCTCGAGATCACCTGATAGGTGTCACCTGAC 691  
 211 LeuAlaProIleAlaThrValLysHisAlaLysSerProGlyThrLys 227  
 692 TTTTGGCCCCAGCGGGTTG ..... GTTAAGGATGTGAATGCACAT 732  
 228 PheLeuLeuLeuProAspMetIleLysGlyLeuPheGlyLysGlu 244  
 733 TTTATC ..... ACCGATTTGTGGATAGTA 758  
 244 uPheLeuTyrGlnThrArgPheLeuArgGlnLeuValIleTyr ..... 258  
 759 AAGCAGCCCGCATATGAAAGGCTGTTTGCATAATGCTGTATACAAAG 808  
 259 ..LeuCysGlyGlnValIleLeuAspGlnIleCysSerAsnIleMetLeu 274  
 809 CCTAGTGGGCGGTAAAGATGTGGATGCCGTG ..... 840  
 275 LeuLeuGlyGlyPheAsnThrAsnAsnMetAsnMetSerArgAlaSerVa 291  
 841 ..... CTGCGTGCACGTAGCGCTAGATGGCGCGCGGAT ..... 873  
 291 ITyrAlaAlaHisThrLeuAlaGlyThrSerValGlnAsnIleLeuHisT 308  
 874 ..... GCCCTGCACGTTATGTCAAGCGTCTCCCAAC ..... 909  
 308 rPserGlnAlaValAsnSerGlyGluLeuArgAlaPheAspTrpLysSer 324  
 ..... 324

```
910 .....GGGCATCAGCGCGATGATCTGCATCGGTGCTAGCTGGGCG 951
325 GluThrLysAsnLeuGluLysCysAsnGlnProThrProValArgTyrAr 341
|||||
952 GAACA..... 957
341 gValArgAspMetThrValProThrAlaMetTrpThrglyGlnAspT 358
||||| ..... ||| :|||
958 .....CCTACCCAGATTTCGTGGGGCAGAGAGATG 988
358 rPLeuSerAsnProGluAspValLysMetLeuLeuSerGluValThrAsn 374
::: :||| :||| :||| :|||
989 AAATTCTTCTGTCCTCCACGCCCTGCGCTGCCAGATGTCATCCCGTG 1038
375 LeuIleTyrHisLys.....AsnIleProGluTrpAlaHisValAspPh 389
::|||::|||::|||::|||::|||
1039 ACAGTGTATGAAGAAACAGGCCCATCTGCCGACCTGAACATGCACAGA 1088
389 eIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGluIleHisL 406
::: :||| :||| :||| :|||
1089 TGTG.....AACAAAGCCATGCCC 1108
406 eu...MetGlnGlnGluGluThrAsnLeuSerGlnIleArgCysGluAla 421
|| :||| :||| :||| :||| :|||
1109 TGTGTGTAAGAACCCGGAAGCCGCGCTGAGCATGCGCCGGATGAGACCG 1158
```



OM of: US-09-578-063-47 to: EST:\* out\_format : pfs  
Date: Jan 30, 2002 11:35 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

-MODEL-frame-p2n.model -DEV-rlp  
-Q=/cgn2\_1/USPSPool/US09578063/runat\_30012002\_092150\_26299/app\_query.fasta\_1.488  
-DB-EST -DBT-fastap -SUFFIX=rest -GAPOP=12.000 -GAPEXT=4.000  
-MIMMARCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000 -GAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DEL0P=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blomsum62 -TRANS-human40.ccd1  
-LIST=45 -DOCL=100 -THR\_SCORE=200 -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTEXT=pfs -NORM=ext -MINLEN=0  
-MALEN=2000000000 -USER=US09578063 @CGN1\_1\_3861 -NCPU=6  
-ICPU=3 -LONGLOG -NO\_XLPHY -WAIT -THREADS=1

## Search information block:

Query: US-09-578-063-47

Query length: 423

Database sequences: 11351937

Database length: 1077921985

Search time (sec): 1532.800000

Score_list:	Strd Orig	ZScore	EScore	Len	! Documentation
gb_hnc:AK019504	+ 1836.00	3644.64	7.6e-194	2927	AK019504 Mus musculus 0 day ne
gb_hnc:AK009571	+ 1139.50	2256.79	1.5e-116	1349	AK009571 Mus musculus adult ma
gb_hnc:AK009875	+ 1139.50	2256.78	1.5e-116	1349	AK009875 Mus musculus adult ma
gb_hnc:AK010093	+ 1139.50	2256.78	1.5e-116	1350	AK010093 Mus musculus adult ma
gb_hnc:AK010139	+ 1139.50	2256.78	1.5e-116	1350	AK010139 Mus musculus adult ma
gb_hnc:AK009431	+ 1139.50	2256.76	1.5e-116	1352	AK009431 Mus musculus adult ma
gb_hnc:AK009479	+ 1138.50	2254.81	2.0e-116	1345	AK009479 Mus musculus adult ma
gb_hnc:AK009523	+ 1138.50	2254.81	2.0e-116	1345	AK009523 Mus musculus adult ma
gb_hnc:AK009729	+ 1138.50	2254.80	2.0e-116	1346	AK009729 Mus musculus adult ma
gb_hnc:AK010116	+ 1138.50	2254.80	2.0e-116	1347	AK010116 Mus musculus adult ma
gb_hnc:AK010058	+ 1138.50	2254.78	2.0e-116	1348	AK010058 Mus musculus adult ma
gb_hnc:AK009300	+ 1138.50	2254.78	2.0e-116	1349	AK009300 Mus musculus adult ma
gb_hnc:AK009474	+ 1138.50	2254.78	2.0e-116	1349	AK009474 Mus musculus adult ma
gb_hnc:AK009546	+ 1138.50	2254.78	2.0e-116	1349	AK009546 Mus musculus adult ma
gb_hnc:AK009573	+ 1138.50	2254.78	2.0e-116	1349	AK009573 Mus musculus adult ma
gb_hnc:AK009525	+ 1138.50	2254.77	2.0e-116	1350	AK009525 Mus musculus adult ma
gb_hnc:AK009773	+ 1138.50	2254.77	2.0e-116	1350	AK009773 Mus musculus adult ma
gb_hnc:AK010061	+ 1138.50	2254.77	2.0e-116	1350	AK010061 Mus musculus adult ma
gb_hnc:AK009428	+ 1138.50	2254.76	2.0e-116	1351	AK009428 Mus musculus adult ma
gb_hnc:AK010019	+ 1138.50	2254.76	2.0e-116	1351	AK010019 Mus musculus adult ma
gb_hnc:AK010035	+ 1138.50	2254.72	2.0e-116	1356	AK010035 Mus musculus adult ma
gb_hnc:AK010203	+ 1136.50	2250.76	3.3e-116	1350	AK010203 Mus musculus adult ma
gb_hnc:AK009413	+ 1135.50	2248.84	4.3e-116	1345	AK009413 Mus musculus adult ma
gb_hnc:AK010123	+ 1135.50	2248.80	4.3e-116	1345	AK010123 Mus musculus adult ma
gb_hnc:AK010236	+ 1134.50	2246.76	5.3e-116	1347	AK010236 Mus musculus adult ma
gb_hnc:AK009990	+ 1134.50	2246.76	5.3e-116	1347	AK009990 Mus musculus adult ma
gb_hnc:AK009537	+ 1133.50	2244.75	7.2e-116	1350	AK009537 Mus musculus adult ma
gb_hnc:AK009359	+ 1131.50	2240.78	9.3e-116	1346	AK009359 Mus musculus adult ma
gb_hnc:AK009544	+ 1131.50	2240.78	9.3e-116	1346	AK009544 Mus musculus adult ma
gb_hnc:AK010103	+ 1130.50	2238.76	1.6e-115	1347	AK010103 Mus musculus adult ma
gb_hnc:AK009560	+ 1129.50	2236.76	2.0e-115	1346	AK009560 Mus musculus adult ma
gb_hnc:AK010026	+ 1129.50	2236.74	2.0e-115	1349	AK010026 Mus musculus adult ma
gb_hnc:AK010231	+ 1128.50	2234.74	2.6e-115	1348	AK010231 Mus musculus adult ma
gb_hnc:AK009437	+ 1127.50	2232.75	3.3e-115	1346	AK009437 Mus musculus adult ma
gb_hnc:AK010036	+ 1125.50	2228.77	5.6e-115	1348	AK010036 Mus musculus adult ma
gb_hnc:AK010148	+ 1124.50	2226.72	7.3e-115	1349	AK010148 Mus musculus adult ma
gb_hnc:AK010106	+ 1121.50	2220.70	1.6e-114	1349	AK010106 Mus musculus adult ma
gb_hnc:AK010173	+ 1120.50	2218.72	2.0e-114	1346	AK010173 Mus musculus adult ma

gb\_hnc:AK010123 + 1110.50 2198.64 2.7e-113 1350 AK010123 Mus musculus adult  
gb\_hnc:AK009923 + 1109.50 2196.19 3.6e-113 1407 AK009923 Mus musculus adult  
gb\_hnc:AK010091 + 1086.50 2150.56 1.3e-110 1347 AK010091 Mus musculus adult  
gb\_hnc:AK010110 + 1075.50 2128.46 2.2e-109 1353 AK010110 Mus musculus adult

seq\_name: gb\_hnc:AK019504

## seq\_documentation block:

LOCUS AK019504 2927 bp mRNA HTC 05-JUL-2001  
DEFINITION Mus musculus 0 day neonate skin cDNA, RIKEN full-length enriched library, clone:4632427C23, full insert sequence.

ACCESSION AK019504

VERSION AK019504.1 GI:12859754

KEYWORDS CAP trapper.

SOURCE Mus musculus (strain:C57BL/6J) 0 day neonate skin cDNA to mRNA, clone:4632427C23.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2927)

Carninci, P., and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Methods in enzymology. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

REFERENCE 2 (bases 1 to 2927)

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome research. 10 (10), 1617-1630 (2000)

MEDLINE 2049374

PUBMED 11042159

REFERENCE 3 (bases 1 to 2927)

Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,

Kono, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,

Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system-384-format

sequencing pipeline with 384 multichannel sequencer

Genome research. 10 (11), 1757-1771 (2000)

MEDLINE 20530913

PUBMED 11078661

REFERENCE 4 (bases 1 to 2927)

The RIKEN Genome Exploration Research Group Phase II Team and the

FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

REFERENCE 5 (bases 1 to 2927)

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,

Arakawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M.,

Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiroaki, T., Horii, F.,

Imotoh, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, D.,

Kojima, K., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T.,

Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y.,

Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H.,

Sasaki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T.,

Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,

Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A.,

Yoshida, K., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.

Direct Submission

Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of

Physical and Chemical Research (RIKEN), Laboratory for Genome

Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

RIKEN Yokohama Institute, 1-7-22 Suenryo-cho, Tsurumi-Ku, Yokohama,

kanagawa 230-0045, Japan (E-mail:genome-res@gsr.riken.go.jp,

URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,

Fax:81-45-503-9216)

Please visit our web site (http://genome.gsc.riken.go.jp/) for

COMMENT



154	.....	154
1303	TCGGGAGAGACTTTGTTAGCAGATCCTGGGGGATTTGCACCATGTTCTGTTT	1352
155	.....	SerTyrAspG1 158
1353	GTTTTATATCTGTGCATAAATCTGTCTTTCCCTTTGTACTATTATGATGA	1402
158	UWETAATATGPhAspLeuProAlaValIleasnPhelIleuGlnLysT	175
1403	AATGGCTAGGTTGACCTTCCAGCGTATAACTTATCTTACAGAAA	1452
175	hG1GyGlnIuLysLLeuTyrTyrValIgiTyrSerGlnIuYthThmet	191
1453	CGGGCCGAGAAAAGGCTCTTATATGCGGCTACTCAGAGGACCCACCATG	1502
192	GlyPhelIeAlaPhSerThrMetProGluLeuAlaGlnLysIleLysMe	208
1503	GGCTTTATATGCATTTTCCACAATGGCCAGAGTAGCTCATTAATCAAAAT	1552
208	ThyTrpAlaIleuAlaProIleAlaThValLysHisAlaLysSerProG	225
1553	GATATTTGGCTTAGCTCCATATGCGACGCTTAAATATCAAGAAGTCTG	1602
225	LYThrLysPheLeuLeuLeuProAspMetMetIleLysG1YleuPheG1	241
1603	GTACCAAAATTTCTGCTGCGCCAGATATGATCATCAAGTATTATTGGC	1652
242	LysLysGluPheLeuTyrGlnThrArpPheLeuArGlnLeuValIleTyr	258
1653	AGACAAAGATTTTATACACAGACTGAGATTTTTCACACAGCTTTTATATTA	1702
258	LleuCySG1GlnValIleLeuAspGlnIleCysSerasnIleMetLeuL	275
1703	CGTTTGGCCAGATGATCTTGACCAAAATCGCAGCACAATCATCTTAC	1752
275	euleuG1YlPheasnThrAsnMetLAsnMetSerArGlaSerVal	291
1753	TCCTGGAGAGTTTAAACAAACAAATATGACATGACAGACAGCAATGTG	1802
292	TyrAlaIleHisThrLeuAlaGlyhrSerValGlnasnIleLeuHisTr	308
1803	TATGTCGCCCATPACCTCTGTGAAAGCTGTGCGCAAAATCTCTCACTG	1852
308	pSerGlnAlaValAsnSerG1YlIleuArGlaPheAspTrpG1SerG	325
1853	GACCAAGCGCGTGAATCTCGGGGAATCTGCTGCTTGCATCGGGAATG	1902
325	IuThrLysAsnLeuG1uLysCysAsnGlnProThrProValArGTrArg	341
1903	AGACCAAAATATCAGGAAATGTATTCAGCCCAATCTATAGGATCAAA	1952
342	ValArAspMetThrValIleProThrIleMetTrpTrnG1YlGlnAspTr	358
1953	GTTTCGAGATATGATGGTCCCAACAGCAATGTGAGATGCGAGGTCAAGACTG	2002
358	PleuSerAsnProGluAspValLysMetLeuLeuSerGluValTrhAsnL	375
2003	GCTTTCAAAATCCAGATGATGAAAAACATTAATCTTGAAATACCAAC	2052
375	euleuLysTrhLysAsnIleProGluTrpAlaHisValAspPheLleTrp	391
2053	TCATCTPACCAAGAAACATCTCTGATGGGCTCAAGTGGATTCATCTGG	2102
392	G1YleuAspAlaProHisIleArgMetTyrAsnGlnIleIleHisLleMetG1	408
2103	GGGCTGGATGCCCTCAGCGTGTTTACAATGAATTAATACATCTGATGAA	2152
408	nGlnGluGluTrhAsnLeuSerGlnG1YlArgCysGluAlaValLeu	423
2153	GCAG.....GACCCCAACCTTTCCCGAGGAACCTGCAAGGCTCAAAATTTG	2195

seq\_name: gb\_hic:AK009571

seq_documentation_block:	1349 bp	mRNA	HMC	05-JUL-2001
LOCUS	AK009571			
DEFINITION	Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310032608, full insert sequence.			
ACCESSION	AK009571			
VERSION	AK009571.1	GI:12844447		
SOURCE	Cap trapper.			
ORGANISM	Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone:11b:RIKEN full-length enriched mouse cDNA library clone:2310032608.			
REFERENCE	Mus musculus			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1 (bases 1 to 1349)			
AUTHORS	Carninci,P. and Hayashizaki,Y.			
TITLE	High-efficiency full-length cDNA cloning			
JOURNAL	Methods in enzymology. 303, 19-44 (1999)			
MEDLINE	99279253			
PUBMED	10349636			
REFERENCE	2 (bases 1 to 1349)			
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.			
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes			
JOURNAL	Genome.research. 10 (10), 1617-1630 (2000)			
MEDLINE	20499374			
PUBMED	11042159			
REFERENCE	3 (bases 1 to 1349)			
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,T., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,M., Sumi,N., Ishi,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujizake,Y., Inoue,K., Togawa,Y., Irawa,M., Ohara,E., Matsubara,Y., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.			
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer			
JOURNAL	Genome research. 10 (11), 1757-1771 (2000)			
MEDLINE	20530913			
PUBMED	11076681			
REFERENCE	4 (bases 1 to 1349)			
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.			
TITLE	Functional annotation of a full-length mouse cDNA collection			
JOURNAL	Nature 409, 685-690 (2001)			
REFERENCE	5 (bases 1 to 1349)			
AUTHORS	Adachi,J., Aizawa,K., Akahira,S., Akiyama,T., Aono,H., Arai,A., Aikawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Harada,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Horii,F., Imotani,K., Ishi,Y., Itoh,M., Irawa,M., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Soabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toyota,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.			
TITLE	Direct Submission			
JOURNAL	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gscc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)			
COMMENT	Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.			
	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to			





```

1164 ACCTGACCTTACTCTGGCGGATGGATGCCCTCCAGAGCTTACATGAG 1213
403 I1e1le1s1leu1met1c1ng1c1ng1u1g1u1r1p1ra1n 413
1214 ATAGTTACCATGATGGCAGACAGACATACAGAAAT 1246

seq_name: gb_htc:AK010093
seq_documentation_block:
LOCUS AK010093 1350 bp mRNA HRC 05-JUL-2001
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched
library, clone:2310067K20, full insert sequence.
ACCESSION AK010093
VERSION AK010093.1 GI:12845297
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
clone:2310067K20.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 1350)
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome research. 10 (10), 1617-1630 (2000)
JOURNAL
MEDLINE 20499374
PUBMED 11042159
TITLE
3 (bases 1 to 1350)
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Kono,H., Akiyama,J., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishize,T., Harada,A.,
Yamamoto,R., Matsunoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Matshiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunaga,S., Kawai,O.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome research. 10 (11), 1757-1771 (2000)
JOURNAL
MEDLINE 20530913
PUBMED 11076861
TITLE
4 (bases 1 to 1350)
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
JOURNAL
MEDLINE 11076861
PUBMED 11076861
TITLE
5 (bases 1 to 1350)
Aachhi,V., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
Arikawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiroaki,T., Horii,F.,
Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
Kojima,Y., Kono,H., Konda,M., Koye,S., Kurihara,C., Matsuyama,T.,
Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shikaki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,T., Takahashi,F.,
Tanaka,T., Teijima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp,
URL:ftp://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)

```



[illegible]

```
COMMENT
KanaGawa 230-0045, Japan (E-mail:genome-res@gsic.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGAGAAGACGATCCAGACCTCCTTTTCTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermoactivated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. cDNA went
through one round of normalization to Rot - 5.0 and subtraction to
Rot - 25.0. Second strand cDNA was prepared with the primer adapter
of sequence [5'.
GAGAGAGACATTCTGAGTTAATTAATTAATCCCCCCCCCCC 3']. cDNA was cleaved
with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
Host: SOLR.

FEATURES
Source          Location/Qualifiers
               1..1350
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /db_xref="MGI:1898377"
                /db_xref="MGI:1914967"
                /clone="2310069P19"
                /sex="male"
                /tissue-type="tongue"
                /clone_lib="RIKEN full-length enriched mouse cDNA library"
                /dev_stage="adult"
                53..1240
                /note="putative"
                /codon_start=1
                /protein_id="BAA8725.1"
                /db_xref="GI:12845372"
                /translation="MMLLLYTVLSAFGAHGFEGKIGPKPNPEANNVSONITWYGP
SEEAEVTEDGYILGVYRIPGRKNSENICKRPAYAIQHLISATWINTLPNNSIA
FIADADGYDVLGSRNGNTSRKNVYSPOSVEMAFSPDMAKYDEAPIDFVORT
GOELIHVHGSGOTTIGFIFASTPALAKIKRYALAPATATVYTESPPKSLIPK
FLKIVIRGNMFMPHNHYLDOLFTEVCSSRELDLCSNALEIFCGFPKKLNVSREFV
YLCHNPACTSTODLFHHRAOLKGLQAVNMGSLOWMLHYNOCKTPPYVVSAMTVPI
AYVNGGDILLADPDVAALLPKPLPNLIHKELLEPTNLDELTIWADAPQEVYNELTYMA
AED"

BASE COUNT      398 a    289 c    283 g    380 t
ORIGIN
alignment_scores:
Quality: 1139.50           Length: 411
Ratio: 3.495              Gaps: 4
Percent Similarity: 79.319 Percent Identity: 52.798

alignment_block:
US-09-578-063-47 x AKO10139 ..
Align seg 1/1 to: AKO10139 from: 1 to: 1350

       7   ArgGINTPRLleValSerHisArgMetGluMetTrpLeuLeuLeuVal 23
       |||:::|||||:::
       35 AGGCACCTGG.....CACATCAAGAGTCGTGGCTCTATTAGTAAAC 72
       |||:::|||||:::
       23 LAlAyrmEtPhgeIn...ArgAsnValAsnSerValHisMetProThrL 39
       ::::|||:::|||||:::
       73 AAGTGTCGTATCGATTGGAGGTGCACATGCCCATTTTGGAAAAACGCG 122
       ::::|||:::|||||:::
       39 ySaLaVaSpPrOcGiLuNgLuTyrcLuVaLaAlaThrgLuAspGlyTyrrII 72
       |||:::|||||:::|||||:::|||||:::
       123 GTCCCAAAAACCCTGACACAACATGATGATGTCAGATGATACCTAC 172
       |||:::|||||:::|||||:::|||||:::
       56 GlnGlyTyrrProcysGlnGluTyrcLuVaLaAlaThrgLuAspGlyTyrrII 72
       |||:::|||||:::|||||:::|||||:::|||||:::
       173 TGGGGAATATCCAAGTAGGAATAATGAAGTTGTACTGAAGATGCGCTACAT 222
```



72 eLeuSerValAsnArgIleProArgIleuValGlnProLysLysThrg 89  
 223 TCTGGGGCTGTATAGAAATCTTATGGGAGAAAATTCGAGAAATTCG 272  
 89 LysArgProValValLeuLeuGlnHISgIleuValGlyGlyAlaSer 105  
 273 GCAAGAGACCTGGTGCATATTGCAGCATGTTGATTGCATCAGCCACA 322  
 106 AsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAs 122  
 323 AACCTGGATTCAATCTGCCAACACAGCCTGCTTCATCTCAGAGA 372  
 122 PALAGLPhAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSera 139  
 373 TGTCTGGTATGATGTGTGCTGGGAAACAGTCGAGGAATTCATGCTCC 422  
 139 rglYsHISLysTrpLeuSerIleAspGlnAspGluPheTrpAlaPheSer 155  
 423 GGAATAATGTATACTATTCAACAGACAGTCAGTGAATCTGGGCTTCAGC 472  
 156 TyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIleLe 172  
 473 TTGATGAATATGGCTAATATGACCTTCACGCCACATAGACTCATTTGT 522  
 172 uGlnLysTrpGlyGlnGluLysIleTyrTrpValGlyTyrSerGlnGlyT 189  
 523 ACAGAAAACGACAGAGAGATACACTATGTTGGTCTCAGTCAAGGCA 572  
 189 hTrpMetGlyPheIleAlaPheSerTrpMetProGluLeuAlaGlnLys 205  
 573 CCACTATCGGCTTTATTCCTTTCTACCATCTGCTGCTGCTCAAAAAA 622  
 206 IleLysMetTyrPheAlaLeuAlaProIleAlaTrpValLysHISAlaLys 222  
 623 ATCAAGAGGTTTATGATTCATTCACCAATCTCAGTGAAGTATACAGA 672  
 222 sSerProGlyTrpLysPheLeuLeuProAspMetIleLysGlyL 239  
 673 AAGCCCTTTAAAAAGATTCATCTTATCTCAAGTCTCTCTCAAGGTGA 722  
 239 eupheGlyLysLysGluPheLeu.....TyrGlnTrpArgPheLeu 252  
 723 TATTGGTATCAAAATGTCTATGCCCAACACTACTTGAATCAATTTCTT 772  
 253 ArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleCys 269  
 773 GGTACGGAGATG.....TGCTCAGGAGGAGCTGCTAGATCTTCTCTG 813  
 269 sSerAsnIleMetLeuLeuGlyGlyPheAsnTrpAsnMetLeuAsn 286  
 814 CAGCAAGCCTTATTCATCTCTGTGATTCAGCAAGAAAACCTTAATG 863  
 286 eTserArgAlaSerValTyrAlaAlaHISTrpLeuAlaGlyTrpSerVal 302  
 864 TGAATTCCTTGTATGTATCTAGGCAATATCCAGCAGCAAGCAATCTACT 913  
 303 GlnAsnIleLeuHISTrpSerGlnAlaValAsnSerGlyGluLeuArgAl 319  
 914 CAAACACCTTTCCACTGGCAGCAGCTGCTAAATCTGGGAAGCTTCAAGC 963  
 319 aPheAspTrpLysSerGluTrpLysAsnLeuGluLysCysAsnGlnProT 336  
 964 CTATTAACGGGAGAGCTCATTACAGAACATCTTACACTACATCAAGAAA 1013  
 336 hTrpValArgTyrArgValArgAspMetTrpValProTrpAlaMetTrp 352  
 1014 CGCCTCCCTACTATGATGTGTGACGACATGCGGCAATGCAATGAGTGTG 1063  
 353 ThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeuLe 369  
 1064 AACGGTGGCCATGACATCTGCTGATGCCCAAGATGTGCAAAATGCTGCT 1113

369 uSerGluValTrpAsnLeuIleTyrHISLysAsnIleProGluTrpAlaH 386  
 1114 TCCCAAACTCCCAACCTTTGTATCCATTAAGAGACCTTCTCCATCAAT 1163  
 386 tSValaAspPheIleTrpGlyLeuAspAlaProHISArgMetTyrAsnGlu 402  
 1164 ACCTGACCTTCATCTGGCGCATGGATGCGCTCAAGAGGTTTACAAATG 1213  
 403 IleIleHISLeuMetGlnGlnGluTrpAsn 413  
 1214 ATACTTACCATGATGCGCAAGAACTACAGAAAT 1246  
 seq\_name: gb\_hlc:AK009431  
 seq\_documentation block: 1352 bp mRNA HTC 05-JUL-2001  
 LOCUS AK009431  
 DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched  
 library, clone:2310021B16, full insert sequence.  
 ACCESSION AK009431  
 VERSION AK009431.1 GI:12844222  
 KEYWORDS CAP trapper.  
 SOURCE Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,  
 clone\_11b:RIKEN full-length enriched mouse cDNA library  
 clone:2310021B16.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 1352)  
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome research. 10 (10), 1617-1630 (2000)  
 JOURNAL 20499374  
 MEDLINE 11042159  
 REFERENCE 2 (bases 1 to 1352)  
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome research. 10 (10), 1617-1630 (2000)  
 JOURNAL 20499374  
 MEDLINE 11042159  
 REFERENCE 3 (bases 1 to 1352)  
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagasaki, S., Sasaki, N., Carninci, P.,  
 Konno, H., Akiyama, J., Nishi, K., Kitsuaki, T., Tashiro, H., Itoh, M.,  
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohata, E., Watanabe, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome research. 10 (11), 1757-1771 (2000)  
 JOURNAL 20530913  
 MEDLINE 11076861  
 REFERENCE 4 (bases 1 to 1352)  
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)  
 JOURNAL 5 (bases 1 to 1352)  
 REFERENCE Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,  
 Arakawa, T., Carninci, P., Fukuda, S., Furumori, Y., Furuno, M.,  
 Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiroaki, T., Horii, F.,  
 Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J.,  
 Kojima, Y., Konno, H., Kouda, M., Koyu, S., Kurihara, C., Matsuyama, T.,  
 Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y.,  
 Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H.,  
 Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T.,  
 Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,  
 Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yanai, A.,  
 Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
 Direct Submission  
 Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of

Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp, URL: <http://genome-gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT  
Please visit our web site (<http://genome-gsc.riken.go.jp/>) for further details.

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence 15' GAGAGAGAGATCCAGAGTAAATTAATCCCCCCCCCC 3'. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

FEATURES  
Source Location/Qualifiers

1..1352  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/db\_xref="MGI:1894047"  
/db\_xref="MGI:1914967"  
/clone="2310021B16"  
/sex="male"  
/tissue\_type="tongue"  
/clone\_idb="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
55..1242  
/note="putative"  
/codon\_start=1  
/protein\_id="BAB26283.1"  
/db\_xref="GI:12844223"  
/translation="MMLLVTSVLSAFGAGLFGKLPKNPEANMNVYSOMITWGY  
SEEEVETEDYILGVYRIPIYKKNSENIIGRPVAYLOHGLIASATMNTLPPNNSLA  
FILADAGDWILSGVIRIPYKKNSENIIGRPVAYLOHGLIASATMNTLPPNNSLA  
GOEKIHYVHGSGTIGTIGTIASTNPALAKTKRYPALAPVATVYTESPFKISLIRK  
VLGNVIFGNKMFPHNYLDOFLGTEVCSRELDLCSNALEIFCGFPKKNLNVSRLDV  
FLGNPAGTSTODLPHMAQLAKSGKLOAYNMGSPLONNLHYNNQTPPYDYVSAMTVPI  
AVNMGHDILADPODVAMLLPKLPLNLYKELIPYHNIDFIWADADPOEYNEIVTMM  
AED"

BASE COUNT 397 a 291 c 285 g 379 t  
ORIGIN

alignment\_scores:  
Quality: 1139.50 Length: 411  
Ratio: 3.495 Gaps: 4  
Percent Similarity: 79.319 Percent Identity: 52.555

alignment\_block:  
US-09-578-063-47 x AK009431 ..

Align seg 1/1 to: AK009431 from: 1 to: 1352

7 ATGGLTTPILLeValSerHisArgMetGluMetTPLeuLeuLeuVala 23  
|||||  
37 AGGCACTGG.....CATTACAGATGTGGCTGCTATTAGTAAC 74  
23 lAlaTyMetPheGln..ArgAsnValaSerValHisMetProThL 39  
|||||  
75 AGGTGCTATCTGCTTTGGAGGTGCACATGCGCTATTGGAAACTGG 124  
39 yAlaValaSerProGluAlaPheMetAsnIleSerGluIleIleGlnHis 55  
|||||  
125 GTCCCAAAACCCTGAGCAACATGAGATGTAAGTATGATTAAGTAC 174

56 GInGlyTyProCysGluGluTyTrGluValAlaThGluAspGlyTrI 72  
|||||  
175 TGGGGATATCCAGATGAGAAATGAACTGTACTGAAAGATGCTACAT 224  
72 eIleuSerValaSerGluIleProArgGlyLeuValGlnProLysLysTh 89  
|||||  
225 TCTGGGGGTCTATAGAAATCTTATGGGAGAAATAATCTGAAATATTCG 274  
89 lYserArgProValaValLeuLeuGlnHisGlyLeuValGlyAlaSer 105  
|||||  
275 GCAAGAGACTGTCGACATATTTCCACAGATGTTGATTCGATCAGCACA 324  
106 AsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAla 122  
|||||  
325 AACTGATTAACAATCTGCCAAACACAGCTGCTTCTATCTACACAGA 374  
122 pAlaGlyPheAspValTrpMetGlyLysSerArgGlyAsnAlaTrpSer 139  
|||||  
375 TCGTGGCTATGATGTGTGGGAGACAGTCGAGGAATACATGGTCCC 424  
139 rGlyHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSer 155  
|||||  
425 GGAAATATGATATCTATTCACCGACTCAGTGAATTCGGCTTCACG 474  
156 TyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIle 172  
|||||  
475 TTGATGAATATGCTAATATATGACTTCACGCCACCATGACTCATTTG 524  
172 uGlnLysThrGlyGlnGluLysLysLeuTyTrValGlyTrpSerGlnGly 189  
|||||  
525 ACAGAAACTGGACAGAGAAATACATATGTTGTTGCTCCAGGGCA 574  
189 hTrhMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLys 205  
|||||  
575 CCCTATCGCTTTATTCCTTTCTACCAATCTCTCTGCTATAAAAA 624  
206 lIleLysMetTyrrPheAlaLeuAlaProIleAlaThrValLysHisAla 222  
|||||  
625 ATCAAGAGGTTTATGATCATAGCTCCAGTGTGTAAGTATACAGA 674  
222 sSerProGlyThrLysPheLeuLeuLeuProAspMetIleLysGlyL 239  
|||||  
675 AAGTCCCTTAAAGATTCATTAATCTTCTTCTTCTTCAAGTGA 724  
239 euPheGlyLysLysGluPheLeu.....TyGlnThrArgPheLeu 252  
|||||  
725 TATTGGTACAAATGTTATCATGCCCCACAACTACTTAATCAATTTCT 774  
253 ArgGlnLeuValIleTyrrLeuGlyGlnValIleLeuAspGlnIleC 269  
|||||  
775 GGATCGGAAGTG.....TGCTCAGCGGAGCTGCTAGACTTCTCG 815  
269 sSerAsnIleMetLeuLeuGlyGlyPheAsnThrAsnAsnMetAsnM 286  
|||||  
816 CAGCAACGCTTATTCATCTCTGTGGATTTCACAGAAACCTTAAAG 865  
286 eSerArgAlaSerValTyrrAlaAlaHisThrLeuAlaGlyThrSerVal 302  
|||||  
866 TGAATGCTTATGATGTGTATAGGCAATATTCAGACAGAAACATCTACT 915  
303 GlnAsnIleLeuHisTrpSerGlnAlaValaSerGlyGluLeuArgAl 319  
|||||  
916 CAAGACCTTTCACAGCGGACAGCTGTGTAATGTCGGAACCTTCAAC 965  
319 aPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnPro 336  
|||||  
966 CTATACCTGGGAGAGTCCATTACAGAACATGTTAACATTCAGAAAA 1015  
336 hProValaArgTyrrArgValaArgAspMetThrValProThrAlaMetTrp 352  
|||||  
1016 CGCCTCCCTACATATGATGTGACGCAATACCGTCCCAATTCAGATGTGG 1065  
353 ThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeu 369

```

TITLE
Journal
Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp)
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGACGAGAAGATCCAGACCTCTTTTCTTTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. cDNA went
through one round of normalization to Rot = 5.0 and subtraction to
Rot = 25.0. Second strand cDNA was prepared with the primer adapter
of sequence [5'
GAGAGCAGACTTCCTCAGATTAAATTAATTAATCCCCCCCCCCC 3']. cDNA was cleaved
with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
Host: SOLR.

FEATURES
Source
Location/Qualifiers
1..1343
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGI:MGI:1896992"
/db_xref="MGI:MGI:1914967"
/cclone="2310022K08"
/sex="male"
/tissue-type="tongue"
/cclone.lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
50..1237
/note="putative"
/codon_start=1
/protein_id="BAB26316.1"
/db_xref="GI:12844304"
/translation="MWLLIVTSVSAFGAGHLGRKLGPKNPEANMVNSOMITFYGYR
SEEEVTEDEGYILGVYRIRPGKKNSENIGRPAYLOHGLASATWNTLPPNLSLA
FILADAGYDVLGNRSNTWSRKNNVYSFDSVEWMAFSDEMATYDPATIDTFVQKT
GOEKIHVYGDSOGTGTFIASTNPDLAKTKRYYAALAPAVATVYTSPFKISLIIPR
FLKLVINPKMKMFHPNYLDQFLGTVCVSRELIDCSNALEIFCGFPKKNNVSRFDVI
YLGINPAGSTNODLFHWAOLAKSGKLDAYVMNGSPLOWNLHYNQTPPYDVSAATVP
AVNMGGHDILLADPODYAMLPRKLPENLYKHETLIPNNHLDFTIMADAQEVYNELVTMM
AED."

BASE COUNT      394 a       289 c       282 g       378 t
ORIGIN
alignment_scores:
    Quality: 1138.50          Length: 411
    Ratio: 3.492              Gaps: 4
Percent Similarity: 79.319   Percent Identity: 52.555

alignment_block:
US-09-578-063-47 x AK009479 ..

Align seg 1/1. to: AK009479 from: 1 to: 1343
7 ARGATGTTTIPLeValSerHisArgMetGluMetTrrPleuLeuValLeuVa 23
||||| | | | | | | | | | | | | | | | | | | | | | | | | |
32 AGGCACCTGG.....CACTACAAGATGTGGCTGCTATTAGTAAAC 69
23 lAlatYrMeCPheGln..ArGaSnYaValSerVallHisMeCProThL 39
::: :: | | | | | | | | | | | | | | | | | | | | | | | |
70 AAGTGTGCATCTGCATTTGGAGGTCACATGCCCTATTGGAAACCTGG 119
39 yshLaValSpProGluAlaPhMeTaSnllLeserUluleeGlnHs 55

```

```

1011 CGCCTCCCACTACTGATGTGTACGCCATTACCSTGGCAATTTCAGTGTGG 1060
353 ThrGlyGlyAlnAspTrpPleUserAasnProGluAspValIlysMetLeude 369
|||||::||| ||::::::::::||| |||||
1061 AACGGTGGGCATCATCTCGTGATGCCCAAGAAGATGCAGATGTGT 1110
369 uSerGluValThrAsnLeuIleTyrlHisLysSnnlleProGluTrpAln 386
- :::: |::::|::::|::::|::::|::::| :::-
1111 TCCCAACACGCCAACCTTCTGTACCATTAAGAGAGATTTCCCTACAATC 1160
386 lvalasphelietrpglyleuasppalaProHlsargmettyransglu 402
|||||::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1161 ACCTGACCTCATCTGTGGATGATGATGCCCTCCCAAGAGGTTTACATGAG 1210
403 llelleHlsleuMetGlncIngJuglUTrasn 413
|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1211 ATAGTTACCATGTATGGCAGAAAGACTTACAGAAAT 1243

seq_name: gb_htc:AK009473

seq_documentation_block:
LOCUS AK009473 1345 bp mRNA HNC 05-JUL-2001
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched
library, clone:3310022H17, full insert sequence.
ACCESSION AK009473
VERSION AK009473
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
clone:3310022H17.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
METHODS In enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
2 (bases 1 to 1345)
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
METHODS In sequencing. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
3 (bases 1 to 1345)
Shibata,K., Itoh,M., Alzawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Sumi.N., Ishii.Y., Nakamura,S., Harada.M., Nishine.T., Harada.A.,
Yamanoto,R., Matsumoto.H., Sakaguchi.S., Ikegami.T., Kashiwagi.K.,
Fujiwaka,S., Inoue,K., Togawa,K., Izawa,M., Matsura.E., Wataniki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Ohara.S., Kawai.J.,
Okazaki.Y., Muramatsu.M., Inoue.Y., Kita,A., and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
METHODS Genome research. 10 (11), 1757-1771 (2000)
MEDLINE 2050913
PUBMED 11076861
4 (bases 1 to 1345)
THE RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM consortium
FUNCTIONAL annotation of a full-length mouse cDNA collection
NATURE 409, 685-690 (2001)
5 (bases 1 to 1345)
Aachawa,T., Alzawa,K., Akahira,S., Akinura,T., Aono,H., Arai,A.,
Arakawa,T., Carninci,P., Fukuda,S., Fukuishiy.Y., Furuno.M.,
Hanagaki,T., Hara,A., Hayatsu.N., Hiramoto,K., Hiroka,T., Horii.F.,
Imotani,K., Ishii.Y., Itoh.M., Izawa,M., Kurihara.C., Matsuyama.T.,
Kojima.Y., Konno.H., Konda.M., Koze,S., Kurihara.C., Matsuyama.T.,
Miyaokaki,A., Nishi.K., Nomura.K., Numazaki.R., Ohno.M., Okazaki.Y.,
Orido,T., Owa.C., Satto.H., Satto.R., Sakai.C., Sakai.K., Sano.H.,

```



seq_documentation_block:	1346 bp	mRNA	HTC	05-Jul-2001
LOCUS	AK0095523			
DEFINITION	Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:23j10028109, full insert sequence.			
ACCESSION	AK0095523			
VERSION	AK0095523.1			
KEYWORDS	GI:12844369			
SOURCE	CAP trapped.			
	Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,			

REFERENCE	1 (bases 1 to 1346)
AUTHORS	Carninci, P. and Hayashizaki, Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Methods in enzymology. 303, 19-44 (1999)
MDLINE	99278253
PUBMED	10349636
REFERENCE	2 (bases 1 to 1346)

Itch, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashibashi, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome research*. 10 (10), 1617-1630 (2000)

REFERENCE  
3 (bases 1 to 1346)  
Shibata, K., Itoh, M., Aikawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,

TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	genome research. 10 (11), 1757-1771 (2000)
MEDLINE	20530913

**AUTHORS**  
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

**TITLE**  
Functional annotation of a full-length mouse cDNA collection

**JOURNAL**  
Nature 409, 685-690 (2001)

**REFERENCE**  
5 (bases 1 to 1346)

**AUTHORS**  
Adachi, J., Altawa, K., Akahira, S., Akinura, T., Aono, H., Arai, A.,

**TITLE**  
DIRECT SUBMISSION  
Submitted (10-JUN-2000) Yoshinobu Hayashizaki, The Institute of  
Journal

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGACAGAGAGAGATCCAGAGACCTCTTTTCTTTTCTTTTCTT-3'], cDNA was prepared by using Trealase thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGACAGACATCTCGACGTATTATTAATATCAACCCCCCCCC-3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

Location/Qualifiers  
1. .1346

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="Caxon:10090"
/db_xref="MGD:MGI:1895759"
/db_xref="MGD:MGI:1914967"
/clone="2310028109"
/sex="male"
/tissue_type="tongue"
/clone_lib="RIKEN_full-length enriched mouse cdna library"
/dev_stage="adult"
53. .1240

```

BASE COUNT	395 a	289 c	283 g	379 t
ORIGIN	.	.	.	.

alignment_scores:	
Quality:	1138.50
Ratio:	3.492
Percent Similarity:	79.319
	Length:
	Gaps:
Percent Identity:	52.555

```
alignment_block:
```

Align seg 1/1 to: AK009523 from: 1 to: 1346

7 ArgGlnrPrIleValSerHisArgMetGluMetTrpLeuLeuIleLeuVa 23  
|||::||| ::::| | | | | : : : : :  
35 AGGCACATGG.....CACTACAAGATGTGGCTGCTATTAGTAAC 72

23 lAlAtYtMeChegIn...ArgAsnValAsnSerValHisMetProThrL 39  
 73 AAGGTCTATCTCATTCATTTGGAGGGCCACATGGCCCTATTTTGGAAAACTGG 122  
 39 ysaIaValAsnProGluIuAlaPheMetAsnIleSerGluIleIleGlnHis 55  
 123 GTCCCAAAAACCCGTGAACCAACATGAAATGTTAGTCACAGTGAATTAATTAC 172  
 56 GlnGlyTyProCysGluGluTyTgIuValAlaIarhGluAspGlyTyTrrI 72  
 173 TGGGATATATCCAAAGTAGGAATATGACAGTGTACTGTAAGATGGCTACAT 222  
 72 eLeuSerValAsnArgIleProArgGlyLeuValGlnProLysTyThg 89  
 223 TCTGTGGGTCTATAGAAATTCCTTAAGGAGAAAAAATCTGGAAATATCG 272  
 89 LysSerArgProValValLeuLeuGlnHisGlyLeuValGlyAlaSer 105  
 273 GCAAGAGACCTGTGGCAATATTGGCAGCATGGTTTGAATTGCATCAGCCACA 322  
 106 AsnTrpIleSerAsnIleuProAsnAsnSerLeuGlyPheIleLeuAlaAs 1222  
 323 AACGGAGATTACAATCTGTCCAAACAACAGCCGTGGCTTCATCTTACACAGA 3722  
 122 PaIaGlyPheAspValTrpMetGlyAsnSerArgIleAsnAlaTrpSerA 139  
 373 TGCTGGCTATGATGTGTGTGGTGGGAGACAGTGGAGGAATACATGCTGCC 4222  
 139 rGlySHisIstYthrLeuSerIleAspGlnAsnGluPheTrpIlaPheSer 155  
 423 GGAHAATAGTATCTATTTCACCAAGCATCTAGTGAATTTGTGGGCTTTCAGC 4722  
 156 TyrAspGluMetAlaArgPheAspPheProAlaValIlaAsnPheIleLe 1722  
 473 TTTGATGAAGAGCCAAATATATGACCTTCCAGCCACATAGACTCAATTGT 5222  
 172 uGlnIstYthrGlyGlnGluLysIleTyTyrValGlyTyTyrSerGlnGlyT 189  
 523 ACAGAAATCTGGACACAGAGAAATACATCATGTGTGTCTACTCTCAGAGGCA 5722  
 189 hTrhTrpMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLys 205  
 573 CCACATATGGGTTTTATGTGCTTTTCTACCAATCTGTCTGTGGCTTAAGAAA 6222  
 206 IleLysMetTyPheAlaIleuAlaProIleAlaThrValLysHisAlaTy 2222  
 623 ATCAAGAGAGTTTATGTCATTAAGTCTCAAGTGTCTACTGTGAAGATACAGA 6722  
 222 sSerProGlyThrLysPheLeuLeuLeuProAspMetIleLysGlyL 239  
 673 AAGTCCCTTAAAAAAGATGTCTACTTATTCCTAAGTTCTTCTCAAGGTGA 7222  
 239 euPheGlyLysLysGluPheLeu...TyrGlnIhTrArgPheLeu 2522  
 723 TATTTGGTAAACAAATGTTCAAGCCCCACAACTACTAGTCAATTTCTT 7722  
 253 ArgGlnLeuValIleTyLeuGlyGluGlnIValIleLeuAspGlnIleCy 269  
 773 GGTACGGAAGTG...TGCTACCGGAGAGCTCTAGATCTTCTCTG 8133  
 269 sSerAsnIleMetLeuLeuLeuGlyLysPheAsnTrhAsnAsnMetAsnM 286  
 814 CAGCAAGCCTTATTCATCTCTGTGTGGATTTGACAAAGAAACCTTAATG 863  
 286 eSerArgAlaSerValTyTAlaAlaHisThrLeuAlaGlyTrpSerVal 302  
 864 TGAGTCCGCTTGATGTGTATCTAAGGCAATAATCCACAGAACATCTACT 9133  
 303 GlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyLysLeuArgAl 319  
 914 CAAGCCTTTTCCATGGGCAAGCTGTGTAAATCTGGGAGAGCTTCAAG 963

[illegible]





seq\_name: gb\_hlc:AK010116

1214 ATAGTTACCATGATGGCAGAGACTPACAGAAAT 1246

seq\_documentation\_block:

LOCUS AK010116 1347 bp mRNA HTC 05-JUL-2001

DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310068005, full insert sequence.

ACCESSION AK010116

VERSION AK010116.1 GI:12845334

KEYWORDS CAP trapper.

SOURCE Mus musculus (Strain: C57BL/6J) adult male tongue cDNA to mRNA, clone\_11b:RIKEN full-length enriched mouse cDNA library clone:2310068005.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1347)

AUTHORS Carninci,P. and Hayashizaki,Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Methods in enzymology. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

2 (bases 1 to 1347)

AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome research. 10 (10), 1617-1630 (2000)

MEDLINE 20499374

PUBMED 11042159

3 (bases 1 to 1347)

AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Sumi,N., Ishi,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,K., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiiwake,S., Inoue,K., Togawa,K., Iwawa,M., Ohara,E., Matshiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuda,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL Genome research. 10 (11), 1757-1771 (2000)

MEDLINE 20530913

PUBMED 11076861

4 (bases 1 to 1347)

AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL  
REFERENCE  
NUMBER  
AUTHORS

Nature 409, 685-690 (2001)  
5 (bases 1 to 1347)

Adechi,J., Aizawa,K., Akahira,S., Akimura,T., Anno,H., Ara,A.,  
Araaka,T., Carninci,P., Fukuda,S., Furukoshi,Y., Furuno,M.,  
Hara,A., Hayashi,N., Hiramoto,K., Hirooka,T., Hori,F.,  
Harasaki,T., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,  
Iomlani,K., Isihara,Y., Itch,M., Iwama,M., Kato,H., Kawai,J.,  
Kojima,Y., Konno,H., Kouma,M., Koza,S., Kuribara,C., Matsuyama,T.,  
Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,  
Okado,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,  
Sasak,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,  
Soabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,  
Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,  
Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

Title  
JOURNAL

Direct Submission (10-Jul-2000) Yoshihide Hayashizaki, The Institute of  
Subjected (10-Jul-2000) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration and Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Shimo-Ogino-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail:genome-res@gscc.riken.go.jp,  
URI:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216)  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
further details.

CDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues. First strand cDNA was primed with a primer  
[5'-GAGGAGAAGACAGCCAGACTCTCTTTTATTTTTTNNN3'] , cDNA was  
prepared by using trehalose thermo-activated reverse transcriptase  
and subsequently enriched for full-length by captrappe. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5'  
GAGAGGAGATTCTCAGTTAAATTAATAATTCGCCCGCCCCCGCCC3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.  
Host: SOLR.

FEATURES  
SOURCE

Location/Qualifiers

1..1347

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/db\_xref="MGD:MGI:1895801"  
/db\_xref="WGS:MGI:1914967"  
/clone="2310068005"  
/sex="male"  
/tissue-type="tongue"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
51..1238

CDS

putative"

/codon\_start=1  
/protein\_id="BAB26711.1"  
/db\_xref="GI:12845335"  
/translation="MWLLLVTSVLSAFGAHGFLGKGPKNPEANNAVSNMTWTPT  
SEEEVLVEDEGYLIGVRIPRGKRNSSENICKRPAYLDQGLIASATWITNLPPNSSLA  
FILDADYGYSOGLTGSGTFSTSRKVNYSPSVVERMAFSPEMATYDDLPATIDFYORTK  
GOEKILHYVGHSOGTTIGCFIAESTPDLAKRKRYALAPAVATYTESPKSLIRK  
FLKIVPRGNMKMFHNHLDOLEGEVSRELDELGNALFIPEGDKPNLNARSFGV  
YLGNPAKTGTODLFPHRAQLASLOANYNGSSLOMLNHNTCPPTYGVSAATVTPI  
AVNWGGHDIIADPDVAMLLFKPLRNLIYHKELIFYNLDFIMAWDAEQEVINEITYMM  
AED"

BASE COUNT 393 a 290 c 284 g 380 t

ORIGIN

alignment\_scores:  
Quality: 1138.50 Length: 411  
Ratio: 3.492 Gaps: 4  
Percent Similarity: 79.319 Percent Identity: 52.555

alignment\_block:  
US-09-578-063-47 x AK010116 ..

Align seg 1/1 to: AK010116 from: 1 to: 1347

```

7  ArgGlnTrpIleValSerHisArgMetGluMetTrpLeuIleLeuVal 23
   |||.....|
33  AGCGATG.....CACTACAGATGGCTGCTATTATGTAAC 70
   |||.....|
23  AlaTyrMetPheGln...ArgAsnValAsnSerValHisMetProThrL 39
   |||.....|
71  AAGTGTCTATCTGCATTGGAGTGCACATGGCTATTGGAAACTCG 120
   |||.....|
39  ySAlaValAspProGluAlaPheMetAsnIleSerGluIleIleGlnHis 55
   |||.....|
121  GTCCCAAAACCTGAGCAAAACATGATGTAGTCAGATGATCAACTTAC 170
   |||.....|
56  GlnGlyTrpProGlyGluIleValAlaIleThrGlnAspGlyTyrI 72
   |||.....|
171  TGGGGATATCCAGTGAAGATATGAATGTTACTGAAGATGGCTACAT 220
   |||.....|
72  eLeuSerValAsnArgIleProArgGlyLeuValGlnProLysThrG 89
   |||.....|
221  TCTGGGGTATATGATTCCTTATGGGAAGAAAATTCGAGAAATTCG 270
   |||.....|
89  LysSerArgProValValLeuLeuGlnHisGlyLeuValGlyGlyAlaSer 105
   |||.....|
271  GCAAGAGACCTGTGGCATATTTCAGCATGTTGATTGCATCGCCACA 320
   |||.....|
106  AsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAs 122
   |||.....|
321  AACTGGATTCACAAATCTGCCAACACAGCCCTTCATTCAGACGA 370
   |||.....|
122  PalaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerA 139
   |||.....|
371  TGCTGGCTATGATGTGTGCTGGGAACACTCGAGGAATACATGAGTCC 420
   |||.....|
139  rGlyHisLysTrpThrLeuSerIleAspGlnAspGluPheTrpAlaPheSer 155
   |||.....|
421  GGAATAATGTAATCTATTACACAGACTCAGTGAATCTGGGCTTACG 470
   |||.....|
156  TyrAspGlnMetAlaArgPheAspLeuProAlaValIleAsnPhelIle 172
   |||.....|
471  TTGATGTAATGGCTAAATATGACCTTCACACCATATGACTTCATCT 520
   |||.....|
172  uGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGly 189
   |||.....|
521  ACGAAACACGACAGACAGAAAGATACACTGTGTTGGTCACTCGAGGCA 570
   |||.....|
189  hrThrMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLys 205
   |||.....|
571  CCACTATCGGTTTATGCTTTCTTACCAATCTGCTCGGCTAAAAAA 620
   |||.....|
206  IleLysMetLysPheAlaLeuAlaProIleAlaThrValLysHisAlaL 222
   |||.....|
621  ATCAAGAGGTTTATGATTAAGTCCAGTGTCTACTGTGAAGATATCAGA 670
   |||.....|
222  sSerProGlyThrLysPheLeuLeuLeuProAspMetIleLysGlyL 239
   |||.....|
671  AACTCCCTTAAAGATTCACCTTATCTTCTTAAGTTCTTCTCAAGGTGA 720
   |||.....|
239  euPheGlyLysLysGluPheLeu.....TyrGlnThrArgPheLeu 252
   |||.....|
721  TATTGTGAACAAGATGTCATGCCCAACACTAGTATCAATTCCT 770
   |||.....|
253  ArgGlnLeuValIleTyrLeuGlyGlnValIleLeuAspGlnIleC 269
   |||.....|
771  GGACGGAAGTG.....TGCCTACGGAGCTGTAGATCTTCTCTG 811
   |||.....|
269  sSerAsnIleMetLeuLeuGlnGlyLysPheAsnThrAsnAsnMetAsn 286
   |||.....|
812  CACAAACGCTTATTCATCTCTGTGATTTGCAAGAAAACCTTAAAG 861
   |||.....|
286  euSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVal 302
   |||.....|

```

```

862  TGAATCGCTTGTGTCTATCTAGGCAATATCCAGCAGGACATCTACT 911
   |||.....|
303  GlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyLeuAlaGAl 319
   |||.....|
912  CAGAACCTTTTCCACTGGGACACGCTTGTAAATCTGGGAAGCTTCAAC 961
   |||.....|
319  aPheAspTrpGlySerGluThrLysAsnLeuGlnLysCysAsnGlnPro 336
   |||.....|
962  CTATTAATCGGGAGATCCATTACAGAACATGTTTACTTACAAATTCAG 1011
   |||.....|
336  hrProValArgTyrArgValArgAspMetThrValProThrAlaMetTrp 352
   |||.....|
1012  CGCTCCCTACTATGATGTGTGACCATGACCGTCCCAATTCGAGTGTGG 1061
   |||.....|
353  ThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeu 369
   |||.....|
1062  AACGGTGGCCATGACATCTGGCTGAGACCCCAAGATCTCCAAATGCTGT 1111
   |||.....|
369  uSerGluValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAlaH 386
   |||.....|
1112  TCCCAAACTCCCAACCTTCTGTACCATAGAGATTCCTCCATACATC 1161
   |||.....|
386  IsValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGlu 402
   |||.....|
1162  ACCTGACCTCATCTGTGGCGATGGATGGCCCTCAAGAGGTTTCAATGAG 1211
   |||.....|
403  IleIleHisLeuMetGlnGlnGluThrAsn 413
   |||.....|
1212  ATAGTTACCATGATGCGCAGAGAACTAACAAGAT 1244
   |||.....|

```

seq\_name: gb\_hnc:AK010058

seq\_documentation\_block: 1348 bp mRNA HTC 05-JUL-2001

LOCUS AK010058 Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2110066C03, full insert sequence.

ACCESSION AK010058

VERSION AK010058.1 GI:12845238

KEYWORDS CAP trapper.

SOURCE Mus musculus (strain: C57BL/6J) adult male tongue cDNA to mRNA, clone: 2110066C03.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS 1 (bases 1 to 1348)

TITLE Carninci, P. and Hayashizaki, Y.

JOURNAL High-efficiency full-length cDNA cloning

METHOD High-efficiency full-length cDNA cloning

PUBMED 99279253

AUTHORS 2 (bases 1 to 1348)

TITLE Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itch, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome research. 10 (10), 1617-1630 (2000)

JOURNAL 20499374

PUBMED 11042159

AUTHORS 3 (bases 1 to 1348)

TITLE Shibata, K., Itch, M., Alzawa, K., Nagao, S., Sasaki, N., Carninci, P.,

Kono, H., Akiyama, J., Nishi, K., Kitanaka, T., Tashiro, H., Itch, M.,

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,

Yamamoto, R., Matsuno, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system - 384-format

sequencing pipeline with 384 multiplexed sequence

JOURNAL 20530913

PUBMED 11076861

REFERENCE 4 (bases 1 to 1348)





Quality: 1138.50 Length: 411  
 Ratio: 3.492 Gaps: 4  
 Percent Similarity: 79.319 Percent Identity: 52.555

alignment\_block:

us-09-578-063-47 x AK009300 ..

Align seg 1/1 to: AK009300 from: 1 to: 1349

```

7  ArgGlnTrrPleValSerHisArgMetGluMetTrpLeuLeuIleLeuVal 23
   |||:|||||
35  AGGCACTGCG.....CACTACAAAGATGCTGCTCTATATAGTAAC 72
23  LalaTrrMetPheGln...ArgAsnValAsnSerValHisMetProHrl 39
   |||:|||||
73  AAGTGTGCTATGTCATTTGGAGGTGCACATGGCTATTTGGAAACTGG 122
39  ysaIaValAspProGluAlaPheMetAsnIleSerGluIleIleGlnHis 55
   |||:|||||
123  GTCCCAAAACCTGAAAGCAAAACATGATTTAGTCAATGATGATCTTAC 172
56  GlnGlyTrrProCysGluGluTrrGluValAlaThrGluAspGlyTrrI 72
   |||:|||||
173  TGGGATATTCAGATGAGAAATGAAATGTTACTGAAAGATGGCTACAT 222
72  eleuSerValAsnArgIleProArgGlyLeuValGlnProLysLysThrG 89
   |||:|||||
223  TCTGGGGGTATAGAAATCTTATGGGAGAAAAATTCAGAAATATGCG 272
89  lySerArgProValIleLeuLeuGlnHisGlyLeuValGlyGlyAlaSer 105
   |||:|||||
273  GCAGAGACCTGTGGCATTTTGCAGCATGCTTTGATGATCAGCCACA 322
106  AsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAla 122
   |||:|||||
323  AACTGATTCACATCTGCCAAACAGCCTGGCTTCATTCAGACAGA 372
122  palagIyPheAspValTrrMetGlyAsnSerArgGlyAsnAlaTrrPera 139
   |||:|||||
373  TGTGTGATTAATGATGTGGCTGGGGAACACTCGGGAATACATGATGCC 422
139  rglYshIstLysThrLeuSerIleAspGlnAspGluPheTrrPalaPheSer 155
   |||:|||||
423  GGAAGAAATGATATCTATTCACACAGCACTGATTCATTCGGCTTACG 472
156  TyrAspGluMetAlaArgPheAspLeuProAlaValIleAsnPheIle 172
   |||:|||||
473  TTTGATGAATGCTAATATGACCTTCAGCCACCATGACTTCATTTGT 522
172  uGlnLysThrGlnGlnLysIleTrrYrrValGlyTrrSerGlnGlyT 189
   |||:|||||
523  ACAGAAACAGTGACAGAGAGATACACTATGTTGGTCACTTCAGGCA 572
189  hrrHmetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLys 205
   |||:|||||
573  CCGCTATCGGTTTATTTGCTTTCTACCAATCTGCTGCTGGCAAAAA 622
206  IleLysMetTrrPheAlaLeuAlaProIleAlaThrYrrValIshIstAla 222
   |||:|||||
623  ATCAAGAGGTTTATGATTCAGTTCAGCTGCTGCTGTAAGTATACAGA 672
222  sSerProGlyThrLysPheLeuLeuLeuProAspMetMetIleLysGlyT 239
   |||:|||||
673  AACTCCCTTTAAAAAGATTTCTACTTATTCCTAAGTTCTCTCAAGGTA 722
239  euPheGlyLysLysGluPheLeu.....TrrGlnTrrArgPheLeu 252
   |||:|||||
723  TATTGTGTAACAAATGCTATGCCCCACACACTTATGATCAATTTCTT 772
253  ArgGlnLeuValIleTrrLysCysGlyGlnValIleLeuAspGlnIleC 269
   |||:|||||
773  GGTACGGAATG.....TGCTCAGCGGAGAGCTGCTATATCTTCTCTG 813

```

```

269  sSerAsnIleMetLeuLeuLeuGlyGlyPheAsnThrAsnAsnMetAsnM 286
   |||:|||||
814  CAGCAAGCCTTTATTCATCTTCCTGTGTGATTTGACAGAAACTTAATG 863
286  eSerArgAlaSerValTrrAlaAlaHisThrLeuAlaGlyThrSerVal 302
   |||:|||||
864  TGAGTCCCTTTGGTGTGTATCTAGGCGCATATCCAGCAGACATCTACT 913
303  GlnAsnIleLeuHisTrrPserGlnAlaValAsnSerGlyLysLeuAla 319
   |||:|||||
914  CAAGACCTTTTCCATGCGGACAGCTTGTAAATCTGGGAAGCTTCAAG 963
319  aPheAspTrrGlySerGlnThrLysAsnLeuGluLysCysAsnGlnPro 336
   |||:|||||
964  CTATTAAGTGGGAGAGTCCATTTACAGACATGTTACACTACAAATCAGAAA 1013
336  hrrProValArgTrrArgValArgAspMetThrValProHrrAlaMetTrr 352
   |||:|||||
1014  CGCCTCCCTACTATGATGTGTACGCCATGACCGTGCATTTGCAATGCTG 1063
353  ThrGlyGlyGlnAspTrrPleuSerAsnProGluAspValLysMetLeu 369
   |||:|||||
1064  AAGCGTGGCCATGATCCTGCGCTGATCCCAAGATGTCGAATGCTGCT 1113
369  uSerGlnValThrAsnLeuIleTrrHisLysAsnIleProGlnTrrPala 386
   |||:|||||
1114  TCCCAAACTCCCAACCTCTGTACCTAAGAGATTTCTCCCTACATC 1163
386  lsaValAspPheIleTrrPglLeuAspAlaProHisArgMetTrrAsnGlu 402
   |||:|||||
1164  ACCGGAATTCATCTGGCGCATGATGCGCTCAAGAGTTTACAAATGAG 1213
403  IleIleHisLeuMetGlnGlnGluGluThrAsn 413
   |||:|||||
1214  ATAGTTACATGATGATGCGAAGACTACAGAAAT 1246

seq_name: gb_hlc:AK009474

seq_documentation_block:
LOCUS      AK009474      1349 bp      mRNA      05-JUL-2001
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched
            library, clone:2310022103, full insert sequence.
ACCESSION  AK009474
VERSION    AK009474.1 GI:12844295
KEYWORDS   CAP trapper.
SOURCE     Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
            clone.lib:RIKEN full-length enriched mouse cDNA library
            clone:2310022103.

ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 1349)
            Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
            Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
            Genome research. 10 (10), 1617-1630 (2000)
JOURNAL    99279253
MEDLINE    10349636
PUBMED     11042159
TITLE      2 (bases 1 to 1349)
            Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
            Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
            Genome research. 10 (10), 1617-1630 (2000)
JOURNAL    20499374
MEDLINE
PUBMED     11042159
REFERENCE  3 (bases 1 to 1349)
            Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
            Kono, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
            Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
            Yamamoto, H., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
            Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
            Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
            Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

```



```

773 GGTACGGAAGTGTG.....TGCTACGGGAGCTGCTAGATCTCTCTG 813
269 sserasnlmetleuleuleuclglypheasntrhshasnmetasnm 286
814 CACCAACGCTTATTCATCTTCTGATTTGACAAAGAAAACCTTAATG 863
286 eteratgalservatlfralalshlrleualaglythseval 302
864 TGAGTCCCTTGTATCTATCTAGGCGATATCCAGCAAGAACCTTACT 913
303 glasnalleuuh1strserglnalavalasnserglyleuauag1 319
914 CAAGACCTTTCCACTGGCAGACAGCTGCTAAATCTGGAGCTTCAAGC 963
319 apheasptpglysergluthrlysnleuglulyscysasnindprot 336
964 CTATTAAGCTGGGAAGTCCATTACAGAACATGTTACACTACATCAGAAA 1013
336 hprvalatgtyrargvalajargasmethrvalprothralametrp 352
1014 CGCCTCCCTACTATGATGTGTCAGCCCTGACCGCTCAATTCAGTGTG 1063
353 ThrGlyGlyGlnasptpLeuSerAsnProGlnAspValylsMetLeu 369
1064 AAGGTGGCCATGACATCTGCTGCTGATCCCAAGATCTCCGATGCTGCT 1113
369 usergluvalthrshleuileuiletyrhlshlysnleleproglutpralsh 386
1114 TCCCAACCTCCCAACCTCTGTAACCATTAAGAGATCTTCCCTACAATC 1163
386 tsvalasphleileitpglyleuaspalaprohshargmettyrasnclu 402
1164 ACCTGAGCTTCATCTGGCGCATGATGCGCTCAAGAGAGTTTACAATGAG 1213
403 lleileh1sleumetglnnglulglutrhasn 413
1214 ATGATTACATGATGAGCAAGACTTAACAGAAAT 1246
seq_name: gb_hlc:AK009546
seq documentation block:
LOCUS AK009546 1349 bp mRNA HTC 05-JUL-2001
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched
library, clone:2310030L19, full insert sequence.
ACCESSION AK009546
VERSION AK009546.1 GI:12844406
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA,
clone:11b:RIKEN full-length enriched mouse cDNA library
clone:2310030L19.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1349)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
2 (bases 1 to 1349)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome research. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
3 (bases 1 to 1349)
AUTHORS Shobara,K., Itoh,M., Alzawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Kono,H., Akiyama,J., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,M.,
Sunb,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,K., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,

```

```

Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL genome research. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
4 (bases 1 to 1349)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5 (bases 1 to 1349)
AUTHORS Aichi,J., Alzawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
Hasegawa,T., Hara,A., Hayatsu,N., Hiramoto,K., Hirose,T., Horii,F.,
Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
Kojima,Y., Kono,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
Sasaki,D., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Tanaka,T., Teijima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTCTT 3'], cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. cDNA went
through one round of normalization to Rot = 5.0 and subtraction to
Rot = 25.0. Second strand cDNA was prepared with the primer adapter
of sequence [5'
GAGAGAGAGATTCGACTTAATTAATTAATTCACCCCCCCCCC 3']. cDNA was cleaved
with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
Host: SOLR.
FEATURES
source Location/Qualifiers
1. 1349
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGI:1893997"
/db_xref="MGI:1914967"
/clone="2310030L19"
/sex="male"
/tissue_type="tongue"
/clone_11b="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
56. 1243
/note="putative"
/codon_start=1
/protein_id="BAB26352.1"
/db_xref="GI:12844407"
/translation="MMLLVTSVLSAFGAGLCEKGLPKRNEAMNVSOMITWGYR
SESEYEVTEDEYILGVIRIPYKGSNENIGRPVAYLOHGLASATNITLNPNSLA
FLADAGYEDWIGNSGRNKNVYSPDSVEFAFSEDEKAKDLPATIDFIVOKT
GOEKTHYHGSGGTIGFIASSTNALAKKIKRPAAPVATVKTSPFKKISLIPK
FLKATIPKRNKMPNPNYIDOLGTGTCRELLDLCSALATIFGCFDKKNKNSFVY
YLGNPAGTSTDDLHMQLKSGKLKATNMGSPQNNLHNTKPTPTDYDSATKPTPI
AVMNGSHIADPDQVAVMLRLPLNPLVHKEILPYNHLDFIWMADAPQEVNEYVTM

```

BASE COUNT 396 a 290 c 284 g 379 t  
ORIGIN

## alignment\_scores:

Quality: 1138.50 Length: 411  
Ratio: 3.492 Gaps: 4  
Percent Similarity: 79.319 Percent Identity: 52.555

## alignment\_block:

us-09-578-063-47 x AK009546 ..

Align seg 1/1 to: AK009546 from: 1 to: 1349

```

7  ArgGlnTrpIleValSerHisArgMetGlnMetTrpLeuLeuVala 23
   |||||
   |||||
   |||||
38 AGGCACTGG.....CACTACAAAGATGGCTGCTATTAAGTAC 75
   |||||
23 LAlaTyrMetPheGln...ArgAsnValAsnSerValHisMetProThrL 39
   |||||
   |||||
   |||||
76 AAGTGTGCTATCTGCATTGGAGTGCACATGGCTATTGGAAACTGG 125
   |||||
39 ysaIaValAspProGlnIaPheMetAsnIleSerGlnIleIleGlnHis 55
   |||||
   |||||
   |||||
126 GTCCCAAAACCTGACCAACCAATGATGTAGTCAGATGATCACTTAC 175
   |||||
56 GlnGlyTyrProCysGlnGluTyrGlnValAlaThrGlnAspGlyTyrL 72
   |||||
   |||||
   |||||
176 TGGGATATCCAGTGAAGATATGAGTTGTTACTGAAGATGGCTACAT 225
   |||||
72 eleUserValAsnArgIleProArgGlyLeuValGlnProLysLysThrG 89
   |||||
   |||||
   |||||
226 TCTGGGGCTATAGATCTTAAATCTTAAAGGAGAAATCTCGAATATACG 275
   |||||
89 lySerArgProValAlaLeuLeuGlnHisGlyLeuValGlyValAsnSer 105
   |||||
   |||||
   |||||
276 GCAAGAGACCTGTCGATATTGTCAGCATGGTTTATGTCATCAGCCACA 325
   |||||
106 AsnTrpIleSerAsnLeuProAsnAsnSerLeuGlyPheIleLeuAlaAs 122
   |||||
326 AACTGATTAACAATCTGCCAACAACAGCTGGCTTCATCTGACAGA 375
   |||||
122 pAlaGlyPheAspValTrpMetGlyAsnSerArgGlyAsnAlaTrpSerA 139
   |||||
   |||||
   |||||
376 TCGTGGCTATGATGTGTGGTGGGACAGTCGAGGAGATACATGCTCC 425
   |||||
139 rGlyHisLysThrLeuSerIleAspGlnAspGluPheTrpAlaPheSer 155
   |||||
   |||||
   |||||
426 GGAATAATGTACTATTACACAGCTCACTGTAATCTGGGCTTCCAGC 475
   |||||
156 TyrAspGlnMetAlaArgPheAspLeuProAlaValIleAsnPheIle 172
   |||||
   |||||
   |||||
476 TTGATGAATGGCTAAATATGACCTTCCAGCCACATAGACTTCATTGT 525
   |||||
172 uGlnLysThrGlyGlnGluLysIleTyrTyrValGlyTyrSerGlnGly 189
   |||||
   |||||
   |||||
526 ACAGAAACTGACAGAAAGAGATACACTATGTGTGTCACCTCAGGGCA 575
   |||||
189 hrThMetGlyPheIleAlaPheSerThrMetProGluLeuAlaGlnLys 205
   |||||
   |||||
   |||||
576 CCACTATCGGTTTATTCCTTTCTTACCAATCCCTGCTGGCTGAAAAA 625
   |||||
206 IleLysMetTyrPheAlaLeuAlaProIleAlaThrValLysHisAlaL 222
   |||||
   |||||
   |||||
626 ATCAAGAGCTTTATGATGACTAGCTCACTTCTACTGTAAGTATACAGA 675
   |||||
222 sSerProGlyThrLysPheLeuLeuLeuProAspMetIleLysGlyL 239
   |||||
   |||||
   |||||
676 AAGTCCCTTAAAAAGATTTCATTATCTTAAGTTCTTCTCAAGGTGA 725
   |||||
239 euPheGlyLysLysGluPheLeu.....TyrGlnThrArgPheLeu 252
   |||||
   |||||
   |||||

```

```

726 TATTGGTAACAAATGTTCATGCCCACTACTATGATCAATTTCTT 775
253 ArgGlnLeuValIleTyrLeuCysGlyGlnValIleLeuAspGlnIleC 269
   |||||
   |||||
   |||||
776 GGTACGGAAGTG.....TGCTACGGAGAGCTGTGATCTTCTGTG 816
   |||||
269 sSerAsnIleMetLeuLeuGlyGlyPheAsnThrAsnAsnMetAsnM 286
   |||||
   |||||
   |||||
817 CAGCAACGCTTATTCATCTCTGTGATTTGACACAGAAAACTTAATG 866
   |||||
286 etSerArgAlaSerValTyrAlaAlaHisThrLeuAlaGlyThrSerVal 302
   |||||
   |||||
   |||||
867 TGAGTGGCTTGATGCTGATCTAGAGGCATATTCAGAGAAACATCTACT 916
   |||||
303 GlnAsnIleLeuHisTrpSerGlnAlaValAsnSerGlyGluLeuArgAl 319
   |||||
   |||||
   |||||
917 CAAGACCTTTTCCACTGGGCACAGCTTCTAAATCTGGGAAGCTTCAAGC 966
   |||||
319 aPheAspTrpGlySerGluThrLysAsnLeuGluLysCysAsnGlnProT 336
   |||||
   |||||
   |||||
967 CTATTAAGTGGGAGAGTCCATTACAGACATGTTACACTACAAATCGAATA 1016
   |||||
336 hrProValArgTyrArgValArgAspMetThrValProThrAlaMetTrp 352
   |||||
   |||||
   |||||
1017 CGCCTCCCTACTATGATGTGTCAAGCATGACCGTCCCAATGACGTGTGG 1066
   |||||
353 ThrGlyGlyGlnAspTrpLeuSerAsnProGluAspValLysMetLeu 369
   |||||
   |||||
   |||||
1067 AACGTGGCCATGACATCTGCTGATCCCAAGATGTGCGCAATGCTCT 1116
   |||||
369 uSerGlnValThrAsnLeuIleTyrHisLysAsnIleProGluTrpAlaH 386
   |||||
   |||||
   |||||
1117 TCCCAACTCCCAACCTCTGTGTACCAAGAGATTTCTTCCCAATTC 1166
   |||||
386 lSValAspPheIleTrpGlyLeuAspAlaProHisArgMetTyrAsnGlu 402
   |||||
   |||||
   |||||
1167 ACCTGACTTATCTGCGCGCATGAGTGGCCCTCAAGAGTTTACATGAG 1216
   |||||
403 IleIleHisLeuMetGlnGlnGluGluThrAsn 413
   |||||
   |||||
   |||||
1217 ATAGTTACCATGATGCGAAGACTAACAGAAAT 1249

```





Gile 199

Seem

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2002, 22:09:19 ; Search time 67.96 Seconds  
(Without alignments) 461.051 Million cell updates/sec

Title: US-09-578-063-47

Perfect score: 2247  
Sequence: 1 MIEFTSRQWIVSHREMMML.....IHLMOEFTNLSGRCFAVL 423

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: /SIDSR/gcgdata/geneseq/AA1980.DAT:\*  
2: /SIDSR/gcgdata/geneseq/AA1981.DAT:\*  
3: /SIDSR/gcgdata/geneseq/AA1982.DAT:\*  
4: /SIDSR/gcgdata/geneseq/AA1983.DAT:\*  
5: /SIDSR/gcgdata/geneseq/AA1984.DAT:\*  
6: /SIDSR/gcgdata/geneseq/AA1985.DAT:\*  
7: /SIDSR/gcgdata/geneseq/AA1986.DAT:\*  
8: /SIDSR/gcgdata/geneseq/AA1987.DAT:\*  
9: /SIDSR/gcgdata/geneseq/AA1988.DAT:\*  
10: /SIDSR/gcgdata/geneseq/AA1989.DAT:\*  
11: /SIDSR/gcgdata/geneseq/AA1990.DAT:\*  
12: /SIDSR/gcgdata/geneseq/AA1991.DAT:\*  
13: /SIDSR/gcgdata/geneseq/AA1992.DAT:\*  
14: /SIDSR/gcgdata/geneseq/AA1993.DAT:\*  
15: /SIDSR/gcgdata/geneseq/AA1994.DAT:\*  
16: /SIDSR/gcgdata/geneseq/AA1995.DAT:\*  
17: /SIDSR/gcgdata/geneseq/AA1996.DAT:\*  
18: /SIDSR/gcgdata/geneseq/AA1997.DAT:\*  
19: /SIDSR/gcgdata/geneseq/AA1998.DAT:\*  
20: /SIDSR/gcgdata/geneseq/AA1999.DAT:\*  
21: /SIDSR/gcgdata/geneseq/AA2000.DAT:\*  
22: /SIDSR/gcgdata/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2247	100.0	423	22	Human TANGO 294.
2	2076	92.4	390	22	Human TANGO 294 ma
3	1289	57.4	399	22	Human shear stress
4	1289	57.4	399	22	Human lysosomal ac
5	1166	51.9	398	14	RGL precursor Or
6	1166	51.9	221	22	Human TANGO 294 ex
7	1156.5	51.5	395	6	Rat lingual lipase
8	1141.5	50.8	379	15	Canine gastric lipase
9	1141.5	50.8	379	17	Dog gastric lipase
10	1141.5	50.8	380	15	Canine gastric lipase
11	1119	49.8	398	7	Sequence of pregas

12	1119	49.8	398	7	AA660658	Sequence of human
13	1119	49.8	398	17	AA609383	Human gastric lipa
14	1119	49.8	398	22	AA660686	Human lipase prote
15	783	34.8	144	22	AA66070	Human TANGO 294 cy
16	691	30.8	233	22	AA661608	Human protein HP03
17	514.5	22.9	410	21	AA62309	Soybean acid triacyl
18	502	22.3	410	21	AA62307	Rice acid triacyl
19	465	20.7	405	21	AA62310	Soybean acid triacyl
20	398	17.7	286	21	AA62304	Corn acid triacyl
21	262	11.7	157	21	AA62303	Catalpa acid triacyl
22	171	7.6	33	22	AA66066	Human TANGO 294 si
23	160.5	7.1	125	21	AA62305	Corn acid triacyl
24	150	6.7	143	21	AA62301	Corn acid triacyl
25	138	6.1	90	21	AA62306	Corn acid triacyl
26	138	6.1	116	21	AA62302	Catalpa acid triacyl
27	128	5.7	125	21	AA66069	Human TANGO 294 tr
28	105	4.7	430	18	AA62234	C. elegans origin
29	105	4.7	430	18	AA61410	Caenorhabditis ele
30	102	4.5	346	18	AA62308	Pyrodicticum sp. es
31	101	4.5	308	17	AA621905	Pseudomonas wiscon
32	100.5	4.5	286	17	AA621903	Pseudomonas wiscon
33	100.5	4.5	286	18	AA622309	P. wisconsinensis m
34	99	4.4	678	22	AA62271	C glutamylum prote
35	98.5	4.4	388	15	AA62124	Acetobacter pasteu
36	98	4.4	471	20	AA608878	Sucrose fatty acid
37	98	4.4	806	20	AA621797	Alteromonas lipase
38	98	4.4	809	20	AA621796	Alteromonas lipase
39	96.5	4.3	1010	21	AA638012	Human secreted pro
40	95	4.2	336	22	AA639139	Human polypeptide
41	94.5	4.2	340	21	AA620361	Arabidopsis thalia
42	94.5	4.2	392	21	AA620360	Arabidopsis thalia
43	94.5	4.2	1280	22	AA64943	Human secreted pro
44	93.5	4.2	484	21	AA645186	Arabidopsis thalia
45	93	4.1	637	22	AA690967	C glutamylum prote

## ALIGNMENTS

RESULT 1	AA66065	AA66065 standard; Protein: 423 AA.
ID	AA66065	
AC	AA66065	
XX		
DT	30-MAR-2001 (first entry)	
XX		
DE	Human TANGO 294.	
XX		
KW	TANGO protein: INTERCEPT protein; neurological disorder;	
KW	central nervous system; focal brain disorder; bipolar affective disorder;	
KW	global diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;	
KW	senile dementia; Huntington's disease; amyotrophic lateral sclerosis;	
KW	Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;	
KW	neuropsychiatric; psychoactive substance use; anxiety.	
XX		
OS	Homo sapiens.	
XX		
PN	W0200077239-A2.	
XX		
PD	21-DEC-2000.	
XX		
PF	24-MAY-2000; 2000WO-US14858.	
XX		
PR	14-JUN-1999; 99US-0333159.	
XX		
PA	(MILL-) MILLENNIUM PHARM INC.	
XX		
PI	McCarthy SA, Fraser CC, Sharp JD, Barnes TM;	
XX		
DR	WPI; 2001-032313/04.	
XX		
DR	N-PSDB; AAF45131, AAF45132.	
XX		



Db 361 daphrmynelhlmggeetnlsgrcav1 390

# RESULT 3

AAB90783 ID AAB90783 standard; Protein: 399 AA.

XX AC AAB90783;

XX DT 15-JUN-2001 (first entry)

XX DE Human shear stress-response protein SEQ ID NO: 66.

XX KW Human; shear stress-response protein; vascular disease;

XX KW arteriosclerosis.

XX OS Homo sapiens.

XX PN MO200125427-A1.

XX PD 12-APR-2001.

XX PF 02-OCT-2000; 2000MO-JP06840.

XX PR 01-OCT-1999; 99JP-0280976.

XX PA (KYOW) KYOMA HAKKO KOGYO KK.

XX PI (NOJIMA) NOJIMA H.

XX PI Nojima H, Yoshisue H, Obayashi M, Ota T, Kawabata A, Sakurada K;

XX PI Kuga T, Sekine S, Nakamura Y, Sugano S;

XX DR WPI: 2001-266308/27.

XX DR N-PSDB; AAH02906.

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

Claim 60; Page 402-404; 678bp; Japanese.

The present invention provides the protein and coding sequences of a number of human shear stress response proteins. These are useful in the diagnosis, treatment and screening of vascular diseases caused by arteriosclerosis, including heart failure, post-PTCA restenosis and hypertension.

Sequence 399 AA;

Query Match 57.4%; Score 1289; DB 22; Length 399;

Best Local Similarity 63.5%; Pred. No. 1.5e-122;

Matches 235; Conservative 55; Mismatches 80; Indels 0; Gaps 0;

40 AVDPAPFAMISITLHOGCPCEYEATVDGTYILSNRIPRLVQPKKGRPVYLHOG 99

28 avdpapfamisitlhogcpceyeatvdgtyilsnrriprrlvqpkkgrpvvylqng 87

100 LVGASNMNINPNNSLGLADAGFDVVMGNSRGNAGRRKKTSLIDODEFAAFYDGM 159

88 lvgasnmninpnnsllgladagfdvvmgnsrgnagrrkktslidodefaafydem 147

160 AFEDLPVAVINFLQKTGQEKIYVGSQGTMGFIASFSTMPLEAOKIKYFALAPIATVK 219

148 afedlpvavinflqktgqekiyvgsqgtmgfiastfmpelakrikmfalfgvasva 207

220 HAKSPGTFELLIDMMIRGLRKKKEFLVQFRLRLQVLYLCOGAVILDOCSNIMLLGCF 279

208 fctspmakigrpdlhldkldgkelfpsatlkwtgltvchvllkelcgnalcflgcf 267

280 NTNNNMNRASVYAAHTAGTGVONLIHWSQAVNSGELRAPDMGSETKMLKCNQDTPPR 339

268 nernlnmrsvdytcthspsagtsvgnmlhwsqavnsgefakfafaagssaknyfhyngsypr 327

OY 340 YRVRDFTVPTAMTGGQDLSPEDYKMLSEVNTLYAKNIPENAHYDFIWLADAPRM 399

Db 328 yrvrdftvptamtgqdlspedykmlsevntlyaknipenahydfiwladaprw1 387

OY 400 YNEIIHIMQO 409

Db 388 ynkliinlmrk 397

# RESULT 4

AAB66061 ID AAB66061 standard; Protein: 399 AA.

XX AC AAB66061;

XX DT 30-MAR-2001 (first entry)

XX DE Human lysosomal acid lipase protein.

XX KW Gene therapy; TANGO protein; INTERCEPT protein; neurological disorder;

XX KW central nervous system; focal brain disorder; bipolar affective disorder;

XX KW global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;

XX KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;

XX KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;

XX KW neuropsychiatric; psychoactive substance use; anxiety.

XX OS Homo sapiens.

XX PN MO200077239-A2.

XX PD 21-DEC-2000.

XX PF 24-MAY-2000; 2000MO-US14858.

XX PR 14-JUN-1999; 99US-0333159.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PI McCarthy SA, Fraser CC, Sharp JD, Barnes TM;

XX DR WPI: 2001-032313/04.

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

Claim 8; Fig 6; 359bp; English.

The present invention relates to TANGO or INTERCEPT proteins and coding sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057, CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding sequences are useful for the treatment of neurological disorders such as central nervous system (CNS) disorders, CNS-related disorders, focal brain disorders, global-diffuse cerebral disorders and other neurological and cerebrovascular disorders. The CNS disorders include Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome, CC autonomic function disorders such as hypertension and sleep disorders, CC neuropsychiatric disorders, psychoactive substance use disorders, CC anxiety, and bipolar affective disorder. The present sequence is a CC sequence used in a sequence homology alignment with the TANGO/INTERCEPT sequences of the present invention.

Sequence 399 AA;

Query Match 57.4%; Score 1289; DB 22; Length 399;

Best Local Similarity 63.5%; Pred. No. 1.5e-122;

Matches 235; Conservative 55; Mismatches 80; Indels 0; Gaps 0;

40 AVDPAPFAMISITLHOGCPCEYEATVDGTYILSNRIPRLVQPKKGRPVYLHOG 99

CC	enzymatic bioconversion, e.g. hydrolysis or transesterification.
XX	
SO	Sequence 398 AA;
Query Match	51.9%; Score 1166; DB 14; Length 398;
Best Local Similarity	54.6%; Pred. No. 4.9e-110;
Matches 219;	Conservative 70; Mismatches 102; Indels 10; Gaps 3;
OY	17 MALLILVAFYMFQRNVSVH-----MPKAVDPEAFMNISETIIOGIPCEEYEVATEDG 70
DB	1 mvlifmvaallsa-igtthgijfgksapt--npevnmhisgmisyswygpkxyevtedg 56
OY	71 YILSVNRIRPRGIVQKKKGSRRVYVLLQGIYVGSASNWTSNIPNNISGLFIADAGDVMWG 130
DB	57 yllenvnrilpygkhsngnrgqfprvflqhgllasasnwlslpnmslatflldadagysvwlq 116
OY	131 NSRGVMSRKHKHTLSIDODEFWAFSYDEMARFEDLPAVINFLIQKTGOEKIYVYGSGQT 190
DB	117 nsrgtwtwrrnlygspdsvefwafsfidemakypdlatridfivkeggelhlvyngsgqt 176
OY	191 MGFIAFSTMPBELAQIKIMYPALAPATYVKAHKSPTGKFLLEPDMWIKGLFGKKEPLVQTR 250
DB	177 lgfiafstmpkiaeikifeyalapatvkyktslnkrlfrrptmfkllfgkklfypnhf 236
OY	251 FLRQIVYILCGVILIDQICSNIMLLGGTNTNNMNSRSVYAATHTLGTSONIILHWSQ 310
DB	237 fdqglatvycsretlnvicsnalflifgfidsanlmsrldvyvshnpatsvqnmllhwq 296
OY	311 AVNSGELRAFMDGSEFTKNLEKCNOPTPVRYRYRDMTVPRTAMTGGODMLSPEDYKMLLS 370
DB	297 aksngnfqafnwgspagnvvhfnqpcpbyynvtamvplavwsgndwlaodpqdvdlilp 356
OY	371 EYTNLDLYHKNIPEMAHVDFTIGLDAPHRMYNEIILHMOOE 411
DB	357 klsnlllyhkellpyhnlhdfiwamnapgeyvelismmakdk 397
RESULT 6	
AAB66068	
ID	AAB66068 standard; Protein: 221 AA.
XX	
AC	AAB66068;
DX	
DT	30-MAR-2001 (first entry)
XX	
DE	Human TANGO 294 extracellular domain.
XX	
KW	TANGO protein; INTERCEPT protein; neurological disorder;
KW	central nervous system; focal brain disorder; bipolar affective disorder;
KW	global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
KW	senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
KW	Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
KW	neuropolychiatic; psychoactive substance use; anxiety.
XX	
OS	Homo sapiens.
XX	
PN	WO200077239-A2.
XX	
PD	21-DEC-2000.
XX	
PF	24-MAY-2000; 2000WO-US14858.
XX	
PR	14-JUN-1999; 99US-033159.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
XX	
PI	McCarthy SA, Fraser CC, Sharp JD, Barnes TM;
XX	
DR	WPI; 2001-032313/04.
XX	
PT	TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for screening assays and diagnostic assays and for the treatment of

PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's  
 PT disease -

PS Claim 8; Page 326; 359pp; English.

CC The present invention relates to TANGO or INTERCEPT proteins and coding  
 CC sequences (see AAF5121-F45136 and AAF45138-F45139 and AAB66031-B66057/  
 CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding  
 CC sequences are useful for the treatment of neurological disorders such as  
 CC central nervous system (CNS) disorders, CNS-related disorders, focal  
 CC brain disorders, global-diffuse cerebral disorders and other  
 CC neurological and cerebrovascular disorders. The CNS disorders include  
 CC Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic  
 CC lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,  
 CC autonomic function disorders such as hypertension and sleep disorders,  
 CC neuropsychiatric disorders, psychoactive substance use disorders,  
 CC anxiety, and bipolar affective disorder.

XX Sequence 221 AA;

Query Match 51.8%; Score 1165; DB 22; Length 221;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-110;  
 Matches 221; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 VHMPTKADPFAFMNISIIQHOGYPCFEYEVATEDGYILSYNRIPRGIVQKKTGSRV 93  
 Db 1 vhmptkavdpfaafmnsiilqhpqpcseyevatedgyilsvnrilprglvqpkktgsrpfv 60  
 QY 94 VLLQGLVGASNMVSNLPPNNSLGFILADAGFDVMNGNSRGNAMSRKHTLSIDDEFWA 153  
 Db 61 vllqglvlgasnmvsnlppnnsilgfildagfdvmngnsrgnawarkhtlsiddqefwa 120  
 QY 154 FSYDEMARFDLPVAVINFLQKTGQEKIYVYSGTGMGFIAFSMPLEAOKIKMFALA 213  
 Db 121 fsydemarfdlpavavinfllqktgqekiyvygsgtgmgfiafstmpelagikmfalala 180  
 QY 214 PIATYKHAKSPTKFLLPDMXIKGLFGKKEFLYQTRFLRQ 254  
 Db 181 piatykhaksptkelllpdmxikglfgkkellyqtrflrq 221

RESULT 7

AA50322 ID AAF50322 standard; protein; 395 AA.

XX AC AAF50322;

XX DT 17-JAN-1992 (first entry)

XX DE Rat lingual lipase protein.

XX KW Lingual lipase; enzyme; EC-3.1.1.3; ss.

XX OS Rattus rattus.

XX GB2142337-A.

XX PD 16-JAN-1985.

XX PF 29-JUN-1984; 84GB-0016581.

XX PR 01-JUL-1983; 83GB-0017989.

XX PR 05-SEP-1983; 83GB-0023759.

XX PA (CELL-) CELTECH LTD.

XX PI Carey NH, Williamson R;

XX DR WPI; 1985-014450/03.

XX DR N-PSDB; AAN50385.

XX PT New lingual lipase protein for treatment of lipase deficiency - also

PT new pre-lingual lipase protein and related products  
 XX Disclosure: Fig 5; 15pp; English.

CC This protein may be expressed in a transformant host organism  
 CC and may be used for the treatment of lipase deficiency.

XX Sequence 395 AA;

Query Match 51.5%; Score 1156.5; DB 6; Length 395;  
 Best Local Similarity 54.7%; Pred. No. 4.5e-109;  
 Matches 216; Conservative 69; Mismatches 109; Indels 1; Gaps 1;

QY 17 MWLLLVAVMFO-RVNVSVHMTKAVDPAFMNISIIQHOGYPCFEYEVATEDGYILSY 75  
 Db 1 mwllllvavmfo-rvnsvshmtkavdpafmnsiilqhpqpcseyevatedgyilsv 60  
 QY 76 NRIPRGIVQKKTGSRPVLLQHLVGASNMVSNLPPNNSLGFILADAGFDVMNGNSRGN 135  
 Db 61 nrilprglvqpkktgsrpfvllqhlvlgasnmvsnlppnnsilgfildagfdvmngnsrgn 120  
 QY 136 AMSRKHKTLSIDDEFWAFSYDEMARFDLPVAVINFLQKTGQEKIYVYSGTGMGFIA 195  
 Db 121 tswrkxnyvpsdsveafsfdemakylpatinfivqktgqekihyvgsgtrtlgfia 180  
 QY 196 FSTMBELAKITMYTALAPATVKAHKSPTKFLLPDMXIKGLFGKKEFLYQTRFLRQ 255  
 Db 181 fstmbelakitktiyalpvatlvkysplkksiflptflkimtgmfkmlphkyidfdfl 240  
 QY 256 VYLGGOVYLDQICSNIMLLGFTNNMNMMSRASVYAAHTLAGTSVONILHMSOAVNSG 315  
 Db 241 grevcsrevyldicentlilfcgfdknlvnsrfdvtyghmpagcsvqdlvmaqlvrsq 300  
 QY 316 ELRARDMGSEETKLNLEKQNPTEVRYRVDMVTVPAMTGGQDWLSNPEDVKLLSEVTNL 375  
 Db 301 kfgatfwgspqgmlyhnoqktkpeydvsaemtyvavawngndlladpqdvamllpklsnl 360  
 QY 376 IYHKNIPENAHDFITGLDAPHRMNEYIHLMOE 410  
 Db 361 lfhekellaynhidflwamdapevynemismnaed 395

RESULT 8

AA56870 ID AAR56870 standard; Protein; 379 AA.

XX AC AAR56870;

XX DT 20-FEB-1995 (first entry)

XX DE Canine gastric lipase.

XX KW Canine gastric lipase; CGL; dog; fat; mucoviscidosis; enzyme;  
 KW bioconversion; exocrine pancreatic insufficiency.

XX OS Canis familiaris.

XX WO9413816-A.

XX PD 23-JUN-1994.

XX PF 16-DEC-1993; 93WO-FR01260.

XX PR 16-DEC-1992; 92FR-0015201.

XX PA (LJOU ) INST RECH JOUVEINAL.

XX PI Benicourt C, Blanchard C, Junien J;

XX DR WPI; 1994-217890/26.

XX DR N-PSDB; AAO68388.



ID	AAFS6871 standard; Protein: 380 AA.
XX	
AC	AAFS6871;
XX	
DT	20-FEB-1995 (first entry)
XX	
DE	Canine gastric lipase (Met-CGL).
XX	
KW	Canine gastric lipase; CGL; dog; fat; mucoviscidosis; enzyme;
KM	bioconversion; exocrine pancreatic insufficiency.
OS	Canis familiaris.
XX	
PN	M09413816-A.
XX	
PD	23-JUN-1994.
XX	
PF	16-DEC-1993; 93MO-FR01260.
XX	
FR	16-DEC-1992; 92FR-0015201.
XX	
PA	(LJOU ) INST RECH JOUVEINAL.
PI	Benicourt C, Blanchard C, Junien J:
XX	
DR	WPI; 1994-217890/26.
XX	
N-	PSDB; AAO68389.
PT	Recombinant canine gastric lipase and nucleic acid encoding it -
XX	
PS	are used for improving absorption of ingested fat, treating
XX	mucoviscidosis etc. and in enzymatic bio-conversions
XX	
PS	Claim 14; Fig 9A; 52pp; French.
CC	The sequence given below is the sequence of figure 9A, altered
CC	according to the amendments described on page 2 of the appended
CC	letter.
CC	CGL is used to improve absorption of ingested fat, in healthy and
CC	sick patients (e.g. having altered levels of gastric lipase); to
CC	treat conditions associated with insufficiency (or lack) of lipases,
CC	esp. mucoviscidosis or exocrine pancreatic insufficiency and partic.
CC	where immobilised, for bioconversions, e.g. hydrolysis or
CC	transesterification (other mammalian gastric lipases, or derivs.,
CC	can be used in this application).
XX	
SQ	Sequence 380 AA;
	Query Match 50.8%; Score 1141.5; DB 15; Length 380;
	Best Local Similarity 56.5%; Pred. No. 1.4e-107;
	Matches 210; Conservative 67; Mismatches 90; Indels 5; Gaps 3
OY	37 PTKANDPAPFNINISIIIOHOGPCPEEYEATBDGTYILSNRIPLGLVOKKTKGSRRVULL 96
Dd	:         :               :      :
Dd	8 pt--npewtmisgmftcywypaaeyevtedgyllgldirpygrknsenigrtpvafl 64
OY	97 QHGVLGASMSINSIPNNISLGFIADAGFDVYMNSRGNAWRKKHTLSIDODEFAFXY 156
Dd	:      :      :      :      :      :      :
Dd	65 qhgllaestwnslpnmslalfiladagdvdlgnsrgrtwarrrnlyspdsvefwaif 124
OY	157 DEMARFEDLPVINFLITQTGOEKITYVGSGQTMGFIAFSTMPELAQIKMYFALAPIA 216
Dd	:      :      :      :      :      :      :
Dd	125 demakydipatidflktkgdqklhyvghbsqgttlglafstnpklrklkfyalapa 184
OY	217 TYKHAKSFGCTREFLLPDMIRKLFEKKETLYTRFLRQ-LVYYLCGGVILLDICSIMLL 275
Dd	:    :    :    :    :    :    :    :    :    :    :
Dd	185 ltkyfeellinklmlypsfklifgnkif-yphhfddglaleavesrevdlcsnalii 243
OY	276 LGGFNTNMNNSRASRYAAHTLAGSYONILHMSOAVNSGELRAPMGSETNLEKNOP 335
Dd	:    :          :    :          :          :      :    :
Dd	244 lcgfctmlmmsrlvdyishnpagtsvqnvlwsgavysgkfqafdwgsprvgmmhynhs 303
OY	336 TPVRVRYVDMDVTPTAMWTGGODMLSPEDVKMLLSEVTNLIVHKNIPEMAHVDFIWLGLDA 395

[illegible]



```

OY 190 TMGFIATSTMPBELAQKIMYFALAPATVNAKSPGTFKLLPDMKIGLFGKKEFLYQT 249
DB 176 tlgfiatstmpslakritfyalapvatvkytkslinklrvpsglfkfifgdklfyphn 235
OY 250 RFLROLVLYLGGVILDOICSNIMLLGGFNTNNMNSRASVYAHAHTLAGTSVONILHMS 309
DB 236 ffdgflatevcsremnlilcsnalficgfsdkntsrldvylshnpagtsvqgmflwt 295
OY 310 QAVNSGELRAFDWSEETKLEKNOPTVRVRYVRDMVPTAMTGGODMLSNPEDVKML 369
DB 296 gavsqgkfqaydwsqpyvgrmhdydgsqpyynvtamvplavwngskdlldpddvglll 355
OY 370 SEVTNLTYHKNIPEMAHVDFTMGDAPHRMNTNEIHLMOOE 411
DB 356 pklpnllyhkeipfynhldfiwamdapevyndivsmisde 397

RESULT 12
AAP60658
ID AAP60658 standard; Protein: 398 AA.
XX
AC AAP60658;
XX
XX 22-AUG-1991 (first entry)
XX
DE Sequence of human pregastric lipase.
XX
KW Cystic fibrosis therapy; enzyme; lipase deficiency.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..19
FT Protein 20..398
XX
PN WO601532-A.
XX
PD 13-MAR-1986.
XX
PF 15-AUG-1985; 85MO-GB00364.
XX
PR 21-AUG-1984; 84GB-0021210.
PR 15-AUG-1985; 85MO-GB00364.
PR 01-JAN-1986; 86GB-0008597.
XX
XX (CELL-) CELLTech LTD.
XX PA (LOWE/) LOWE P A.
XX
XX PI Lowe PA;
XX
XX WPI; 1986-081634/12.
XX DR N-PSDB; AAN60566.
XX
XX PT New gastric lipase protein, esp. of human origin - for treating
XX lipase deficiency, and DNA sequences coding for it
XX
XX PS Disclosure; Fig 3; 39pp; English.
XX
XX CC The inventors claim a pregastric lipase protein and a gene encoding
XX it. Gastric lipase protein is useful for oral administration to
XX CC treat lipase deficiency, e.g. cystic fibrosis or pancreatitis.
XX
XX SQ Sequence 398 AA;

Query Match 49.8%; Score 1119; DB 7; Length 398;
Best Local Similarity 52.2%; Pred. No. 3e-105;
Matches 210; Conservative 73; Mismatches 107; Indels 12; Gaps 2;
OY 17 MWLLILVAYMFORNVSMHPTKAV-----DPEAFMISEITIOHGRCCEYEYVATD 69
DB 1 mwllilmasl-----svlgthglfghklhpgspvevmlsqmiltlywypneeyevtd 55

```

```

OY 70 GYILSVNRIIPRGVLQPKKTSRPPVLLQHLVGASNMISLNPNNISGLFIADAGFDVWM 129
DB 56 gylilsvnrilprgvlqpkktsrgppvllqhlvgasnmislnpnnisglfiadagfdvwm 115
OY 130 GNSGNMNSRKHKLSIDODEFNAFSDMARFPLPANINFILOKTGEEKIYYVGSQGT 189
DB 116 gnsrgnltwarrnlylpspsvelfwafstdeamkylpalditlvkktgqkqilyvghsgt 175
OY 190 TMGFIATSTMPBELAQKIMYFALAPATVNAKSPGTFKLLPDMKIGLFGKKEFLYQT 249
DB 176 tlgfiatstmpslakritfyalapvatvkytkslinklrvpsglfkfifgdklfyphn 235
OY 250 RFLROLVLYLGGVILDOICSNIMLLGGFNTNNMNSRASVYAHAHTLAGTSVONILHMS 309
DB 236 ffdgflatevcsremnlilcsnalficgfsdkntsrldvylshnpagtsvqgmflwt 295
OY 310 QAVNSGELRAFDWSEETKLEKNOPTVRVRYVRDMVPTAMTGGODMLSNPEDVKML 369
DB 296 gavsqgkfqaydwsqpyvgrmhdydgsqpyynvtamvplavwngskdlldpddvglll 355
OY 370 SEVTNLTYHKNIPEMAHVDFTMGDAPHRMNTNEIHLMOOE 411
DB 356 pklpnllyhkeipfynhldfiwamdapevyndivsmisde 397

RESULT 13
AAM09383
ID AAM09383 standard; Protein: 398 AA.
XX
AC AAM09383;
XX
XX 19-AUG-1997 (first entry)
XX
DE Human gastric lipase protein sequence.
XX
XX
KW Duodenal; gastric; lipase; transgenic; plant; recombinant; extract;
KW food; absorption; fat; pancreatic; cystic fibrosis; exocrine; dairy;
KW hydrolysis; trans-esterification; substrate; enzyme; biofuel; human.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..19
FT Protein /note="signal peptide"
FT Protein 20..999
FT Protein /note="mature protein"
XX
XX PN WO9633277-A2.
XX
XX PD 24-OCT-1996.
XX
XX PF 19-APR-1996; 96MO-FR00606.
XX
XX PR 20-APR-1995; 95FR-0004754.
XX
XX PA (BIOC-) BIOCEM SA.
XX PA (LJOV) INST RECH JOUVEINAL.
XX
XX PI Baudino S, Benicourt C, Cudrey C, Gruber V, Lenee P;
XX PI Merot B;
XX
XX DR WPI; 1996-485783/48.
XX DR N-PSDB; AAT58916.
XX
XX PT Recombinant human or canine pre-duodenal lipase prodn. in transgenic
XX plants - useful for facilitating absorption of fat, as bio-catalysts
XX and for prodn. of bio-fuel
XX
XX PS Claim 3; Fig 5; 130pp; French.
XX
XX CC This is the amino acid sequence of the human pre-duodenal (i.e. gastric)
XX lipase enzyme. The sequence can be used to generate transgenic plants

```

CC Producing recombinant lipase in an enzymatically active form.  
 CC Alternatively amino acids 20-23 or 20-73 (1-4 or 1-54 of the mature  
 CC protein, respectively) can be deleted to form the derivatives designated  
 CC delta-4 or delta-54 respectively. Plants, or their extracts, expressing  
 CC the lipases or the truncated derivatives, can be used:  
 CC (a) as pharmaceuticals or food to facilitate absorption of fat, either  
 CC in healthy subjects or in patients with inadequate levels of fat, either  
 CC gastric/pancreatic lipase e.g. those with cystic fibrosis or exocrine  
 CC pancreatic insufficiency, in the elderly or in patients undergoing medical  
 CC treatment which alters fat adsorption; (b) for performing industrial or  
 CC agricultural reactions, e.g. in processing of fats or in the dairy  
 CC industry, for hydrolysis or transesterification reactions, etc., where  
 CC the plant material may provide both enzyme and substrate. The transgenic  
 CC plants can also be used for biofuel production.

XX  
 XX Sequence 398 AA;

Query Match 49.8%; Score 1119; DB 17; Length 398;  
 Best Local Similarity 52.2%; Pred. No. 3e-105;  
 Matches 210; Conservative 73; Mismatches 107; Indels 12; Gaps 2;

QY 17 MMLILVAVMQRVNSVHMPKAV-----DPEAFMNISEIIHOGYPCXEYEVATED 69  
 DB 1 mwilltmasll-----svlgtlthglfgklhpgspevtmnlsmqtlwygypneeyevlved 55  
 QY 70 GYLISVNRIPRGIVQPKTGSRPVLLDHGLVGAASNMISLPPNNSLGFILADAGFDVMM 129  
 DB 56 gyllevnrilpygkksngntqgrpvrlqhgllasaatwislmpnslaflladagydwl 115  
 QY 130 GNSRGNMRSRKHKTLSDIDDEFMAFSYDEMAREFLPAVINILKGTGOKIYYGYSOGT 189  
 DB 116 gnsrgntwarrnllyspdsvefwafsdemakypaldidivkkgqqlhyvghsgqt 175  
 QY 190 TMGFIAFTMPBELAQIKMYAPALPIATVKAHSGPTGKFLPDMITGKGRKEFLYQT 249  
 DB 176 tlgflafstmpslakrlktftalavpavtkslinklirvpqslkfflfgdklfyphn 235  
 QY 250 RFLRQVLYICGOVILDOICSNIMLLGTFNTNMMNSRASVYAHLTAGTSVONILHMS 309  
 DB 236 ffdgflatevcsremnlncsnalffllcgfasknfntarlvytlshnpagtsvqmfhw 295  
 QY 310 QAVNSGELARADGSEKTKLEKQNPFRVRYRVDMYPTAMTGGGOMLSNPEDVYKML 369  
 DB 296 qavnsqkfgaydwsqpvgrmhbydsqppyyrvlamvplavwngkdlldpdpvgvll 355  
 QY 370 SEVNTLIYHKNIPEMAHDFIMGLDAPHRMYNEIHLMOQEE 411  
 DB 356 pklpnlllyhkeipfnhldifwamdapevynldvsmlseck 397

RESULT 14  
 AAB66086  
 ID AAB66086 standard; Protein: 398 AA.  
 XX  
 AC AAB66086;  
 XX  
 DT 30-MAR-2001 (first entry)  
 XX  
 DE Human lipase protein.  
 XX  
 XX Gene therapy; TANGO protein; INTERCEPT protein; neurological disorder;  
 XX central nervous system; focal brain disorder; bipolar affective disorder;  
 XX global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;  
 XX senile dementia; Huntington's disease; amyotrophic lateral sclerosis;  
 XX Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;  
 XX neuropsychiatric; psychoactive substance use; anxiety.  
 OS Homo sapiens.  
 XX  
 XX W0200077239-A2.  
 XX  
 XX 21-DEC-2000.

XX  
 XX 24-MAY-2000; 2000MO-US14858.  
 PF  
 XX  
 XX 14-JUN-1999; 9905-0333159.  
 PR  
 XX  
 XX (MILL-) MILLENNIUM PHARM INC.  
 PA  
 XX  
 XX McCarthy SA, Fraser CC, Sharp JD, Barnes TM;  
 PI  
 XX  
 XX WPI; 2001-032313/04.  
 DR  
 XX  
 XX TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for  
 PT screening assays and diagnostic assays and for the treatment of  
 PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's  
 PT disease -  
 PS  
 XX  
 XX Disclosure; Fig 6; 359pp; English.

CC The present invention relates to TANGO or INTERCEPT proteins and coding  
 CC sequences (see AAF45121-FA5136 and AAF45138-FA5139 and AAB66031-B66057,  
 CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding  
 CC sequences are useful for the treatment of neurological disorders such as  
 CC central nervous system (CNS) disorders, CNS-related disorders, focal  
 CC brain disorders, global-diffuse cerebral disorders and other  
 CC neurological and cerebrovascular disorders. The CNS disorders include  
 CC Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic  
 CC lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,  
 CC autonomic function disorders such as hypertension and sleep disorders,  
 CC neuropsychiatric disorders, psychoactive substance use disorders,  
 CC anxiety, and bipolar affective disorder. The present sequence is a  
 CC sequence used in a sequence homology alignment with the TANGO/INTERCEPT  
 CC sequences of the present invention.

XX  
 XX Sequence 398 AA;

Query Match 49.8%; Score 1119; DB 22; Length 398;  
 Best Local Similarity 52.2%; Pred. No. 3e-105;  
 Matches 210; Conservative 73; Mismatches 107; Indels 12; Gaps 2;

QY 17 MMLILVAVMQRVNSVHMPKAV-----DPEAFMNISEIIHOGYPCXEYEVATED 69  
 DB 1 mwilltmasll-----svlgtlthglfgklhpgspevtmnlsmqtlwygypneeyevlved 55  
 QY 70 GYLISVNRIPRGIVQPKTGSRPVLLDHGLVGAASNMISLPPNNSLGFILADAGFDVMM 129  
 DB 56 gyllevnrilpygkksngntqgrpvrlqhgllasaatwislmpnslaflladagydwl 115  
 QY 130 GNSRGNMRSRKHKTLSDIDDEFMAFSYDEMAREFLPAVINILKGTGOKIYYGYSOGT 189  
 DB 116 gnsrgntwarrnllyspdsvefwafsdemakypaldidivkkgqqlhyvghsgqt 175  
 QY 190 TMGFIAFTMPBELAQIKMYAPALPIATVKAHSGPTGKFLPDMITGKGRKEFLYQT 249  
 DB 176 tlgflafstmpslakrlktftalavpavtkslinklirvpqslkfflfgdklfyphn 235  
 QY 250 RFLRQVLYICGOVILDOICSNIMLLGTFNTNMMNSRASVYAHLTAGTSVONILHMS 309  
 DB 236 ffdgflatevcsremnlncsnalffllcgfasknfntarlvytlshnpagtsvqmfhw 295  
 QY 310 QAVNSGELARADGSEKTKLEKQNPFRVRYRVDMYPTAMTGGGOMLSNPEDVYKML 369  
 DB 296 qavnsqkfgaydwsqpvgrmhbydsqppyyrvlamvplavwngkdlldpdpvgvll 355  
 QY 370 SEVNTLIYHKNIPEMAHDFIMGLDAPHRMYNEIHLMOQEE 411  
 DB 356 pklpnlllyhkeipfnhldifwamdapevynldvsmlseck 397

RESULT 15  
 AAB66070  
 ID AAB66070 standard; Protein: 144 AA.  
 XX  
 XX

AC AAB6070;  
 XX  
 DT 30-MAR-2001 (first entry)  
 XX  
 DE Human TANGO 294 cytoplasmic domain.  
 XX  
 KW TANGO protein; INTERCEPT protein; neurological disorder;  
 KW central nervous system; focal brain disorder; bipolar affective disorder;  
 KW global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;  
 KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;  
 KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;  
 KW neuropsychiatric; psychoactive substance use; anxiety.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO20007239-A2.  
 XX  
 PD 21-DEC-2000.  
 XX  
 PF 24-MAY-2000; 2000WO-US14858.  
 XX  
 PR 14-JUN-1999; 99US-0333159.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI McCarthy SA, Fraser CC, Sharp JD, Barnes TM;  
 DR WPI; 2001-032313/04.  
 XX  
 PT TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for  
 PT screening assays and diagnostic assays and for the treatment of  
 PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's  
 PT disease -  
 XX  
 PS Claim 8; Page 327; 359pp; English.  
 XX  
 CC The present invention relates to TANGO or INTERCEPT proteins and coding  
 CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,  
 CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding  
 CC sequences are useful for the treatment of neurological disorders such as  
 CC central nervous system (CNS) disorders, CNS-related disorders, focal  
 CC brain disorders, global-diffuse cerebral disorders and other  
 CC neurological and cerebrovascular disorders. The CNS disorders include  
 CC Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic  
 CC lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,  
 CC autonomic function disorders such as hypertension and sleep disorders,  
 CC neuropsychiatric disorders, psychoactive substance use disorders,  
 CC anxiety, and bipolar affective disorder.  
 CC  
 XX  
 SQ Sequence 144 AA;  
 SQ

Query Match 34.8%; Score 783; DB 22; Length 144;  
 Best Local Similarity 100.0%; Pred. No. 1e-71;  
 Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 280 NTNNMNSRSASVYAHTLAQSVONILHMSQAVNSGELRAFDWGSSETKNLEKCNQPTPVR 339  
 DB 1 ntmmmsrsasyaahlaqsvqnllhwsqavnsgelelrafdwgselknleknqptpr 60  
 OY 340 YRVADMTVPFAMWTGGODWLSNPEDVKMLSEVTNLIYHNKIPBMAHVDPIWGLDAPHRM 399  
 DB 61 YRVADMTVPFAMWTGGODWLSNPEDVKMLSEVTNLIYHNKIPBMAHVDPIWGLDAPHRM 120  
 OY 400 YNEIHLMOQETNLSGRCFAVL 423  
 DB 121 YNEIHLMOQETNLSGRCFAVL 144



DB 9 NPEANNISOMITWYWGPCOEYVEDGYILGVYRIPHGKNNSENIGKRPVYVLLQHLI 68  
QY 102 GGASNMISINPNNSLGFILLADAGEFDVWGNRGNAMSRKKTLSIDODEFMAVSDEMAR 161  
DB 69 ASATNMIANIPNNSLAFMLADAGIDVWLGNSRGNTMSRKRVVYSPDSVEERMAVSFDEMAK 128  
QY 162 FDLPAVINFLIQTGQEKIYVGYSGTGTGFIASFSTMPBLAOKIKMYFALADIAATVKA 221  
DB 129 YDLPATINFLVQGTGQEKIYVGHOSGTTIGFIASFSTNPLAKIKITFYALAVATVKKY 168  
QY 222 KSPGTFFLLPDMKIGLBEKKEFLVOTRPLROLVYLICQVILLDICSIMILLGCFNT 261  
DB 189 QSPBKRSIFPTFLFKMFCKMFLPPTYEDDLGTEVCSEVLDLCSNTLPIFCGFDR 248  
QY 282 NNNMRSASYAAHTLAGTSVONILHMSQAVNSGELRAFPMGSETKNLEKCNOPTRVRYR 341  
DB 249 KNLVSRFEDVYLGHNPAGISVODFLHMAQLVRSGKFOAFWGSPOMLHYNOKTPREYD 308  
QY 342 VRDTPVTAMTGGODMLSNPDEVKMLSEVTMLIYHKNIPEMAHYDFIWLDAPIHRMYN 401  
DB 309 VSAMTVPVAVWNGNDILADPDQVAMLLPKLSNLFHKEILAVNHLDPIWAMDAPQEVYN 368  
QY 402 EIIHMQOE 410  
DB 369 EMISMAED 377

## RESULT 2

US-09-073-674-17  
Sequence 17, Application US/09073674  
Patent No. 5998189  
GENERAL INFORMATION:  
APPLICANT: Blanchard, Claire  
APPLICANT: Junien, Jean-Louis  
TITLE OF INVENTION: Recombinant Dog Gastric Lipase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Warner-Lambert Company  
STREET: 2800 Plymouth Road  
CITY: Ann Arbor  
STATE: Michigan  
COUNTRY: U.S.A.  
ZIP: 48105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/073,674  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Crissey, Todd M.  
REGISTRATION NUMBER: 37,807  
REFERENCE/DOCKET NUMBER: 5072-DI-66-TWC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 734 622-7530  
TELEFAX: 734 622-1553  
TELEX:  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 377 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-073-674-17

Query Match 51.7%; Score 1161; DB 2; Length 377;  
Best Local Similarity 57.7%; Pred. No. 2e-114;

Matches 213; Conservative 63; Mismatches 93; Indels 0; Gaps 0;  
QY 42 DPEAFNISEIIOHOGYPCGEYEVAREDDGYILSVNRPBGLVQPKTGSPVYVLLQHLI 101  
DB 9 NPEANNISOMITWYWGPCOEYVEDGYILGVYRIPHGKNNSENIGKRPVYVLLQHLI 68  
QY 102 GGASNMISINPNNSLGFILLADAGEFDVWGNRGNAMSRKKTLSIDODEFMAVSDEMAR 161  
DB 69 ASATNMIANIPNNSLAFMLADAGIDVWLGNSRGNTMSRKRVVYSPDSVEERMAVSFDEMAK 128  
QY 162 FDLPAVINFLIQTGQEKIYVGYSGTGTGFIASFSTMPBLAOKIKMYFALADIAATVKA 221  
DB 129 YDLPATINFLVQGTGQEKIYVGHOSGTTIGFIASFSTNPLAKIKITFYALAVATVKKY 168  
QY 222 KSPGTFFLLPDMKIGLBEKKEFLVOTRPLROLVYLICQVILLDICSIMILLGCFNT 261  
DB 189 QSPBKRSIFPTFLFKMFCKMFLPPTYEDDLGTEVCSEVLDLCSNTLPIFCGFDR 248  
QY 282 NNNMRSASYAAHTLAGTSVONILHMSQAVNSGELRAFPMGSETKNLEKCNOPTRVRYR 341  
DB 249 KNLVSRFEDVYLGHNPAGISVODFLHMAQLVRSGKFOAFWGSPOMLHYNOKTPREYD 308  
QY 342 VRDTPVTAMTGGODMLSNPDEVKMLSEVTMLIYHKNIPEMAHYDFIWLDAPIHRMYN 401  
DB 309 VSAMTVPVAVWNGNDILADPDQVAMLLPKLSNLFHKEILAVNHLDPIWAMDAPQEVYN 368  
QY 402 EIIHMQOE 410  
DB 369 EMISMAED 377

## RESULT 3

US-08-227-108-18  
Sequence 18, Application US/08227108  
Patent No. 5807726  
GENERAL INFORMATION:  
APPLICANT: Blanchard, Claire  
APPLICANT: Junien, Jean-Louis  
TITLE OF INVENTION: Recombinant Dog Gastric Lipase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/227,108  
FILING DATE: 03-APR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fanucci, Allan A.  
REGISTRATION NUMBER: 30,256  
REFERENCE/DOCKET NUMBER: 7620-033  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 379 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-227-108-18

```

1 SOFTWARE: PatentIn Release #1.0, Version #1.25
2
3 CURRENT APPLICATION DATA:
4 APPLICATION NUMBER: US/09/073,674
5
6 FILING DATE:
7
8 CLASSIFICATION:
9
10 ATTORNEY/AGENT INFORMATION:
11 NAME: Clissey, Todd M.
12 REGISTRATION NUMBER: 37,807
13 REFERENCE/DOCKET NUMBER: 5072-D1-66-TMC
14 TELECOMMUNICATION INFORMATION:
15 TELEPHONE: 734 622-7530
16 TELEFAX: 734 622-1553
17
18 TELEX:
19
20 INFORMATION FOR SEQ ID NO: 18:
21
22 SEQUENCE CHARACTERISTICS:
23 LENGTH: 379 amino acids
24 TYPE: amino acid
25

```

TITLE OF INVENTION: Recombinant Dog Gastric Lipase  
 NUMBER OF SEQUENCES: 21  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Pennie & Edmonds  
 STREET: 115 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: U.S.A.  
 ZIP: 10036  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/227,108  
 FILING DATE: 03-APR-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Fanucci, Allan A.  
 REGISTRATION NUMBER: 30,256  
 REFERENCE/DOCKET NUMBER: 7620-033  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212 790-9090  
 TELEFAX: 212 869-8864/9741  
 TELEX: 66141 PENNIE

```

; INFORMATION FOR SEQ ID NO: 3
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 379 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
US-08-227-108-3

```

Query Match	50.8%	Score 1141.5	DB 1	Length 379
Best Local Similarity	56.5%	Pred. No. 2,4e-112		
Matches 210	Conservative 67	Mismatches 90	Indels 5	Gaps 3

QY	37	PTKAVDEAFENISEIIIOHOGYPCPEEYVATEDDYISLVNRIPGLVQPKKTSRPVLL	96
Dd	7	PT---NEVYINISQMTTMYGYPRAEEYVATEDDYIIGIDIRIPGRKNSENIGRRPAFL	63
QY	97	QHGLVGASAMWISLPPNNSLGFILLADAGFVMMGNSRGNAWSRKHKILSLIDODEFMAFSY	156
Dd	64	QHGLLASATWISLPPNNSLAFILLADAGDYVDLGNSSRGNTMARNNLYSPDSVEFMAFSF	123
QY	157	DEAMRFDPVAVINITLQKTGOEKIYYGYISOGTTMGCTIASTIMBELAQKIKMTFALAPLA	216
Dd	124	DEAMKYDLPALIDELIKRTGQDKLHHYGHSGGTTIGTIASTINPKLAKRIKTEYALAPVA	183
QY	217	TVKAKSPGCTFFLLPDMWIKGLGKKKEFLYQTRFELRO-LVIYICGQVILIIOJCSNMLL	275
Dd	184	TVKTTETFLNKLMVPSFLEKFLIRGNKIF-YPHHFFQOFLATEYCSRETYDLCSNAFLTI	242
QY	276	LGGFENTNMNMSRASVYAAMHTLACTSVONILHMSOAVNSELRAFDGSEFTKMLEKQNP	335
Dd	243	ICGDFDTNMLNMSRLDYVLSHPRACTSVQVNLHMSQAVKSKFOAFDMGSPSVQNMHHQS	302
QY	336	TPVRYRVARDMTVPFAAMTGGODWLSNPEDYKMLLSEVTNLITRKNIIDEMAHVDFIWGLDA	395
Dd	303	MPRYNLTDMHVPFAVWNGNDLLADHDVOLLSTKLPNLITYHRKIPRYNHLDFIWAMDA	362
QY	396	PHRYMNETIHLM 407	
Dd	363	FOAYINELIVSM 374	

RESULT 6  
US-09-073-674-3  
Sequence 3, Application US/09073674  
Patent No. 5998189  
GENERAL INFORMATION:  
APPLICANT: Blanchard, Claire  
APPLICANT: Beaucourt, Claude  
APPLICANT: Junten, Jean-Louis  
TITLE OF INVENTION: Recombinant Dog Gastric Lipase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Warner-Lambert Company  
STREET: 2800 Plymouth Road  
CITY: Ann Arbor  
STATE: Michigan  
COUNTRY: U.S.A.  
ZIP: 48105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/073.674  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Clissey, Todd M.  
REGISTRATION NUMBER: 37,807  
REFERENCE/DOCKET NUMBER: 50772-D1-66-TMC  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 734 622-7530  
TELEFAX: 734 622-1553  
TEXT:  
INFORMATION FOR SEQ ID NO:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 379 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-073-674-3

Query Match	50.8%;	Score	1141.5;	DB	2;	Length	379;
Best Local Similarity	56.5%;	Pred. No.	2.4e-112;				
Matches	210;	Conservative	67;	Mismatches	90;	Indels	5;
						Gaps	3;

```

QY      37  PTKAVNDEAFANISEIIOHOGYPCEEYEVAETEDQYILISVRNIPRGVLQPKKTSRPAVLL 96
Dh      7  PT---NEVYMNISQMITYMGYPAEEYEVAETEDQYILIGRIDRIPRGKNSENIGRRPAVL 63
QY      97  OHGLVGASAWISNLPPNNSLGFILADAGFDVWNGNSNGNMRKHKTLSTIDODEFAFST 156
Dh      64  OHGLLASATWISNLPNNSLAFILADAGDYWLCLNSGNWMAWNNLYYSPDSVEFWAFSP 1233
QY      157  DEMARFDLPVNIINILKOTQOEKYYUYGSGOTMGRIASFSTMBELAQIKMEFALAPIA 216
Dh      124  DEMAKYDLPATIDELIKKTQODKILHYVGHSGOTTIGTIAFSTNKLKRTKTYALAPVA 1833
QY      217  TVKRAKSPGTFFELLPPMMIKLGLEGGKEFYQTFRLQ-LVIYLCQVILIDQCSNIMLL 275
Dh      184  TVKTTETELKLMLVPSFELKRLITGNKIF-YPHHFQFQFLATEVCSRETYDLLCSNMLFI 242
QY      276  LGGFNTNNMNSRASVYAHAHTLACTSVONILHMSOAVNSGELEAFDWSGSETKMLEKCNOP 335
Dh      243  ICGFDITNMLNSRLDYVLSHNPATSVONVLIHMSQAVKSGFQAFDWSGSPVQMMHMYHQS 3020
QY      336  TPVRIYVRDXTVPPAAKMTGGODMLSNEDYKMLSEVTNLIYHKNIPEMAHNVDFIAGDA 3355
Dh      303  MPYYNLTDMVYPAVWVNGNDLADPHVDLLSKLPIYHKKRIPRYNHDFIWAMDA 3620
QY      396  PHRYNTEIHL 407
Dh      363  POAVYNTEIWSM 374

```

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525





```

? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/227,108
? FILING DATE: 03-APR-1994
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: FANUCCI, Allan A.
? REGISTRATION NUMBER: 30,256
? REFERENCE/DOCKET NUMBER: 7620-033
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 212 790-9090
? TELEFAX: 212 869-8864/9741
? TELEX: 66141 PENNIE
? INFORMATION FOR SEQ ID NO: 16:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 380 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? US-08-227-108-16

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

APPLICANT: GAVIN, Kimberly
TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,105
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard Aron
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 430 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-105-22

```

```

Query Match 4.7%; Score 105; DB 1; Length 430;
Best Local Similarity 21.9%; Pred. No. 0.013;
Matches 61; Conservative 42; Mismatches 87; Indels 88; Gaps 13;

```

```

Qy 85 PKKTGRPVYLLQGLVGGASNMISNLPNN-----SLGF-----ILADAGFDVW--MGN 131
Db 25 PEKESR-----QKTKGKEMASRNLSNLEEDLQGFEDETYSMAQSALENFMGK 78
Qy 132 S-----RGNAMSRKH-----TLSDIDEFMAFSYDENARFDLPVINFILQKGOEKIY 181
Db 79 SASERMMNAKSRGRAGNGNTEIEEDDEISNATDFTKCDLPGLRNYITRKDKNTEFEK 138
Qy 182 YVGSQGTMGFIASFTEPDLAQIKMYFALAPITVYKHAASPGTKFLLPDMIKGLFG 241
Db 139 RLEHLADNDFG-----KMKLYLAG-----FNILLHGVGS 168
Qy 242 KKEFLYOTRFLQOLVYLCGOVILDOICSNIMLLGFGNTN---NMNRSASVYAHTLA 298
Db 169 KRDLV--TEFENELSDYIMRYVDARKDGLNKKVLLGAINENMKLNCKNVRG----- 217
Qy 299 GTSVGNILHWSQAV---NSGEL-----RAFDWGSE 325
Db 218 ---OSTISWARSIRKKNMSOOLLILIDNIEAPDWRSD 251

```

```

RESULT 12
US-08-484-106-22
Sequence 22, Application US/09484106
Patent No. 5614518
GENERAL INFORMATION:
APPLICANT: STILLMAN, Bruce
APPLICANT: ZELT, Stephen P
APPLICANT: KOBAYASHI, Rumi I
APPLICANT: RINE, Jasper
APPLICANT: JESS, Mark T
APPLICANT: GUNALY, Francis J
APPLICANT: LAURINSON, Patricia

```

```

APPLICANT: HERSKOWITZ, Ira
APPLICANT: LI, Joachim J
APPLICANT: GAVIN, Kimberly
TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,106
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard Aron
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 430 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-106-22

```

```

Query Match 4.7%; Score 105; DB 1; Length 430;
Best Local Similarity 21.9%; Pred. No. 0.013;
Matches 61; Conservative 42; Mismatches 87; Indels 88; Gaps 13;

```

```

Qy 85 PKKTGRPVYLLQGLVGGASNMISNLPNN-----SLGF-----ILADAGFDVW--MGN 131
Db 25 PEKESR-----QKTKGKEMASRNLSNLEEDLQGFEDETYSMAQSALENFMGK 78
Qy 132 S-----RGNAMSRKH-----TLSDIDEFMAFSYDENARFDLPVINFILQKGOEKIY 181
Db 79 SASERMMNAKSRGRAGNGNTEIEEDDEISNATDFTKCDLPGLRNYITRKDKNTEFEK 138
Qy 182 YVGSQGTMGFIASFTEPDLAQIKMYFALAPITVYKHAASPGTKFLLPDMIKGLFG 241
Db 139 RLEHLADNDFG-----KMKLYLAG-----FNILLHGVGS 168
Qy 242 KKEFLYOTRFLQOLVYLCGOVILDOICSNIMLLGFGNTN---NMNRSASVYAHTLA 298
Db 169 KRDLV--TEFENELSDYIMRYVDARKDGLNKKVLLGAINENMKLNCKNVRG----- 217
Qy 299 GTSVGNILHWSQAV---NSGEL-----RAFDWGSE 325
Db 218 ---OSTISWARSIRKKNMSOOLLILIDNIEAPDWRSD 251

```

```

RESULT 13
US-08-602-359A-34
Sequence 34, Application US/08602359A
Patent No. 5942430
GENERAL INFORMATION:
APPLICANT: ROBERTSON, Daniel E.
APPLICANT: MURPHY, Dennis
APPLICANT: REID, John
APPLICANT: MAFFIA, Anthony
APPLICANT: LINK, Steven

```

RESULT 14  
US-08-232-519-2  
Sequence 2, Application US/08232519  
Patent No. 5484725  
GENERAL INFORMATION:  
APPLICANT: Kageyama, Bunji  
APPLICANT: Nakae, Masanori  
TITLE OF INVENTION: NORBORANE TYPE ESTER HYDROLASE  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Neave  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA

QY	38	TKADPEFMNISIIOHOCPEEYE	-VATEGGYILSVNRIPRG--LVOPKKTGR--	91		
Db	84	TPDYDIEAFI-----	-KNFHAENPDAAATEBASGERKQVYVGEHTLVNRDVGQOEG	1353		
QY	92	-PVVLLGHVGGASNMISLNPNSLGFLLADAGFDVYMGNSRCNANSRRH--	-KTLSID	147		
Db	136	TRIVLV--HGFGGDISNMVL--	-LTQDAAERRVIAFDL----	PGHGASSKNVGGTLAF--	1868	
QY	148	QDEFMASYSDEMAMFDLPAYINFLQKTGOEKIYVIGSOGTIMGFLAFSTMPRLAOKIK	207			
Db	187	-----LAGVSELDLQTLIEKAVVHESLIG--	-GGJATLTLRLDRPDQVA	2277		
QY	208	MYFLAPATATYKNAKSPQTKFLLLPDMIMKGLFEKKEFLQTRPELRDVLVY--	-LGGQVI	2644		
Db	228	SLNLLAPRGL--	-GKDVNADEFI-----	SAPVDSRSSDKMKVADLVQVILNKALVGRKM	2767	
QY	265	LDQICSNIMLLDGFNTNNMMSRASVYAAHTLGTAVONILHWSQAVNSGGEIARAEDMGS	3244			
Db	277	VDAY-----	-LRARRLDGARD--	-ALHVIKACFPN-----	-GHQADDLHSLVLAGA	3177
QY	325	ETKLEKCNQPTPVRYRVRDQTVPTAMTGGQDMLSNPDEVKILISEVTMLIYHK--	-NIP	3622		
Db	318	ET-----	-PTQIFWGGEDEILSVNSNAAGLPDVI	PVTUYEETGHLR	3566	
QY	383	EMAHVDFWIGLDRHRYNNEIHL--	-MOQEEFNISGCRSEA	421		
Db	357	QLEHATDV-----	-NKAIATLVVDPEALASMAEMDA	386		

```

RESULT 15
US-08-456-956-2
Sequence 2, Application US/08456956
Patent No. 5686285
GENERAL INFORMATION:
APPLICANT: Kaheyama, Bunji
APPLICANT: Nakae, Masanori
APPLICANT: Yagi, Shigeo
TITLE OF INVENTION: No. 5686285Dornane Type Ester Hydrolase
NUMBER OF SEQUENCES: 2

```

Thu Jan 31 07:54:02 2002

us-09-578-063-47.ra1

Page 9

Search completed: January 30, 2002, 23:01:20  
Job time: 3001 sec

CORRESPONDENCE ADDRESS:  
ADDRESS: Fish & Wildlife

ADDRESS: Fish & Neave  
STREET: 1251 Avenue of the Americas

CITY: New York  
STATE: New York

COUNTRY: USA  
ZIP: 10020

COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS

SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/456,956

FILING DATE: 01-JUN-  
 CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-96286

FILING DATE: 22-APR-1993  
PRIOR APPLICATION DATA:..

APPLICATION NUMBER: US 08/232,519  
FILING DATE: 22-APR-1994

ATTORNEY/AGENT INFORMATION: 22 APR 1994

NAME: HALEY JR, James F  
REGISTRATION NUMBER: 27,794

REFERENCE/DOCKET NUMBER: SHGN-4 DIV 1  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090

TELEX: 14-8367  
INFORMATION FOR SEQ ID NO: 3

SEQUENCE CHARACTERISTICS:

LENGTH: 388 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
 MOLECULE TYPE: protein

8-456-956-2

Query Match	4.48; Score 98.5; DB 1; Length 388;
-------------	-------------------------------------

Best Local Similarity 20.2%; Pred. No. 0.052;  
Matches 81; Conservative 64; Mismatches 142; Indels 113; Gaps 20;

38 TKAVDPEAFMNISETIOHOGYPCEEYE--VATEDGYITSVNRIPRG--LVOPKKTGSR-- 91

```

      | |||:      : : | |||      : : | : : ||:
84 TPDVDTFAFI-----KTHAENBODAAATFEDASAGEBKROVTVGEHTI NVRDVGSTGEG 13

```

[illegible]

92 -P-V-L-D-H-G-L-V-G-A-S-I-N-J-I-M-P-E-N-N-S-L-O-F-I-L-A-D-A-G-E-D-Y-W-M-K-N-S-R-U-N-A-M-S-R-I-N-H--K-I-T-S-I-D-14  
|::|| | |||: | ::| ||: | : | |: |:  
-----

136 TPILV-HGFGDISNML--LTQDALAERRVIAFDL-----PGHGASSKNVGTGLAF- 18

```

148 QDEFWAFSYDEMARFDLPVAVINFILOKTGQEKIYVGYSGQTTMGFAFSTMPELAQKTK 200
      | :: :: | | : || | | || : :

```

b 187 -----LAGVSELLQTKIEKAHVGHSLG---GGIALTLRLRDPQVA 22

208 MYFALARIATVKHAKSPGTEKLLLPDMIKGLFGKKEFLQTRFLRQVLIY---LCGYI 266

228 SLNLLAPAGL---GKDVNADEI-----SAFVDSSESRDMKAVLQMLVYNKALVGKKM 27

265 LDQICSNIMLLGGFNTNNMNSRASVYAAHTLACTSVQNIHLWSQAVNSGELRAFDNGS 322

```

277 VDAY-----LRARRDGDARD--ALHVIKACFPN-----GHQADDLHSLVLAGA 31

```

325 ETKNLEKCNQPTPVRRYRVKMDVPTAMWTGGQDWLSNPEDVKMLLSEVTNLIYHK--NIP 388

318 EM-----PTQIFMGKEDEILSVSNAGLPDVPYIPVTVEETGHP 35

383 EWAHDEIWI.DAPHNYNETIHI.-MOOETNI.SOGCEA 421

357 O'FARROW-----WATARI ENKORDEBARI SAKABARA 306

[illegible]

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2002, 22:13:59 ; Search time 57.04 Seconds

(without alignments)  
564.899 Million cell updates/sec

Title: US-09-578-063-47

Perfect score: 2247  
Sequence: 1 MLETLISQWIVSHRMEMML.....IHLMOQRETNLSQRCRAVL 423

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.68:\*  
2: PIR.1:\*  
3: PIR.2:\*  
4: PIR.4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1290	57.4	399	2	S41408 lysosomal acid lip
2	1290	57.4	399	2	G01416 lysosomal acid lip
3	1164.5	51.8	395	1	LIRTT triacylglycerol 1i
4	1139	50.7	397	1	JC4017 triacylglycerol 1i
5	1119	49.8	398	2	S07145 triacylglycerol 1i
6	752.5	33.5	403	2	T33198 hypothetical prote
7	751	33.4	411	2	T22290 hypothetical prote
8	730	32.5	405	2	T22675 hypothetical prote
9	575.5	29.2	426	2	T20480 egg-specific prote
10	574.5	28.6	559	2	JT0949 triglyceride lipas
11	505	22.5	443	2	T39540 triglyceride lipas
12	478.5	21.3	1585	2	T31611 hypothetical prote
13	475	21.1	467	2	T41053 triglyceride lipas
14	445	19.8	344	2	E84526 probable lysosomal
15	422.5	18.8	548	2	S37969 probable triacygl
16	412.5	18.4	460	2	T39443 probable triglycer
17	411.5	18.3	413	2	T43170 probable triacygl
18	294	13.1	431	2	D86318 protein P15H18.6 l
19	285	12.7	538	2	S6842 probable membrane
20	271.5	12.1	573	2	S64754 probable membrane
21	234.5	10.4	509	2	G96766 protein lipase F2P
22	149	6.6	62	2	S59904 lipase, pregastric
23	143	6.4	336	2	F83425 probable esterase/
24	126	5.6	987	2	A64474 hypothetical prote
25	113	5.0	412	2	B72391 conserved hypothet
26	112.5	5.0	40	2	S19539 triacylglycerol 1i
27	105	4.7	430	2	T23001 hypothetical prote
28	103.5	4.6	798	2	F66714 probable protease
29	101.5	4.5	929	2	JH0262 Quer protein - Eme

30	101	4.5	436	2	JC5021 platelet-activatin
31	99.5	4.4	286	2	D75217 probable 2-acetyl-
32	99.5	4.4	1280	2	B34087 hypothetical prote
33	99	4.4	460	2	G36764 unknown protein F2
34	97.5	4.3	322	2	B70957 probable epna prot
35	97.5	4.3	456	2	E69391 hypothetical prote
36	96.5	4.3	287	1	E64053 hypothetical prote
37	96.5	4.3	1275	2	I38588 reverse transcript
38	96	4.3	809	2	S67153 probable membrane
39	96	4.3	934	2	B82364 DNA polymerase I V
40	95.5	4.3	1259	4	GHRUL1 retrovirus-related
41	95	4.2	973	2	T41201 isoleucyl-trna syn
42	94.5	4.2	561	2	G83913 hypothetical prote
43	93	4.1	1275	2	B38096 line-1 protein ORF
44	93	4.1	2311	1	TYCHSR kinase-related pro
45	92.5	4.1	1275	2	S65824 reverse transcript

## ALIGNMENTS

```

RESULT 1
S41408
lysosomal acid lipase (EC 3.1.1.-) / sterol esterase (EC 3.1.1.13) precursor - human
C:Species: Homo sapiens (map)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 18-Jun-1999
C:Accession: S41408; M39315; S47187
R:Amels, D.; Merkel, M.; Eckerskorn, C.; Greten, H.
Eur. J. Biochem. 219, 905-914, 1994
A:Title: Purification, characterization and molecular cloning of human hepatic lysoso
A:Reference number: S41408; MUID:94155897
A:Accession: S41408
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-399 <AMP>
A:Cross-references: EMBL:X76488; NID:9434305; PIDN:CAA54026.1; PID:9434306
R:Anderson, R.A.; Sando, G.N
J. Biol. Chem. 266, 22479-22484, 1991
A:Title: Cloning and expression of cDNA encoding human lysosomal acid lipase/choleste
A:Reference number: A39315; MUID:92042192
A:Accession: A39315
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-15, 'P', 17-399 <AND>
A:Cross-references: GB:M74775; NID:9187151; PIDN:AAA59519.1; PID:9187152
R:Du, H.; Gregory, G.A.
submitted to the EMBL Data Library, April 1994
A:Description: Structural conservation of putative functional motifs between mouse an
A:Reference number: S47187
A:Accession: S47187
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-22, 'R', 24-399 <DUH>
A:Cross-references: EMBL:Z31690; NID:9506430; PIDN:CAA63495.1; PID:9506431
C:Genetics:
A:Gene: GDB:LIPA
A:Cross-references: GDB:120153; OMIM:278000
A:Map position: 10q24-10q25
A:Superfamily: triacylglycerol lipase, lingual
C:Keywords: carboxylic ester hydrolase; glycoprotein

Query Match 57.4%; Score 1290; DB 2; Length 399;
Best Local Similarity 61.0%; Pred. No. 6.1e-106;
Matches 242; Conservative 59; Mismatches 94; Indels 2; Gaps 1;

OY 15 MEMMLILVAYNMFQNNVSVHMPK--AVDPRAFNNISEIIHQHGPCEEEVATEDEYI 72
I:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 MKMRFLGLVAVLWTLVLSSESGGKLTAVDPTNNVSEIISYWGFPSEELVETEDGYI 60
OY 73 LSVNRIPIRGIVQPKRTGSRPVLLDHGLVGASNNISNLPPNNSIGFTILADAGFDVMKNS 132
I:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 LCLNRIPIGRKHSXDKGKRPVYFLDHGLLADSSNNVYTNLANSLSGFTILADAGFDVMKNS 120

```



R:Timmermans,M.Y.J.; Reekmans, G.; Teuchy, H.J.H.; Kuipers,L.P.M.  
Biochem. J. 314, 931-936, 1996  
A:Title: Inhibition studies on calf pregastric esterase: the enzyme has no functional th  
A:Reference number: S64678; MUID:96177869  
A:Accession: S64678  
A:Molecule type: protein  
A:Residues: 24-40/248-253 <TIM>  
C:Comment: Pregastric esterase is a major fat-digesting enzyme.  
C:Genetics:  
A:Gene: pge  
C:Superfamily: triacylglycerol lipase, lingual  
C:Keywords: blocked amino end; carboxylic ester hydrolase; glycoprotein  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-39/Product: pregastric esterase #status predicted <MAT>  
F:33,270/326/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:17/Active site: Ser #status predicted

Query Match 50.7%; Score 1139; DB 1; Length 397;  
Best Local Similarity 53.7%; Pred. No. 1,3e-92;  
Matches 212; Conservative 66; Mismatches 113; Indels 2; Gaps 2;

18 MLLIIIVYVNFQRRVNVSVHMFPTKAVDPKAFNINISEIIHQGYPCPEEYVATEDGYILSVNR 77  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
3 WLLVYVCFIMSGNAFECFLGKIANKPESAMVSCQISITWGYSEMKKVTADGYILQVYR 62  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
78 IPRGLVQPKTKGSRPVVLLQGLVGGASNMWISNLPNNLSGLFIADAGFDVMGNSRGNV 137  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
63 IPRGNANANHLGGRPVYELDGLLGISATNMWISNLPKNSLGLFIADAGVYVWIGNSRGNW 122  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
138 SRHKRLSIDODEFMAFSYDEMAFDPVAVNFIQKQGEKIIYGYSGGTMTGTAFS 137  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
123 AOEHLVYSPSPSPFMAFSPEMAEYDLPSTIDFLIRTGQKRLHYVGHSGGTIGFIAS 182  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
198 TPELAKIKITMYALALAIYAVYKHAKSPGTFLLPDMIKGFLKKEFYQTFRLQ-LV 256  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
183 TSFTLEAKIVFVALAVAVYKYTKSLPKNLALPHFLPKLIFGDMF-YPHTFLEQFLG 241  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
257 IYLCGOVILDOICSNIMLLGGFNTNNMNSRASYAAHTLGTQSVONTLHMSQAVNSGE 316  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
242 VEMCSRETLVDLCKNALFATIGVDNKNFNMNSRLDYIAHNPAQTSVONTLHMRAQVKSCK 301  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
317 LRAFDMGSEFKNLEKQNPVRYRVKDMVYPTAMTGGQDPLSNPEYKMLSEYTNLI 376  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
302 FQAFDMGAPYQONTLMHYQPPPIYNTLTAMNPPIAVWSADNDLADPOQDVDFLSKLSNLI 361  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
377 YHKINPEMAHVDFTMGDLADPHRYNNEIHLMOOEE 411  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
362 YHKEIPNYNHLDFIWMADAPQEVNTEIVSLMAEDK 396  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 5  
S07145  
triacylglycerol lipase (EC 3.1.1.3) precursor, gastric - human  
C:Species: Homo sapiens (man)  
C:Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 18-Jun-1999  
C:Accession: S07145; S27102; S04942  
R:Boedmer, M.W.; Angel, S.; Yarranton, G.T.; Harris, T.J.R.; Lyons, A.; King, D.J.; Plect  
Biochim. Biophys. Acta 909, 237-244, 1987  
A:Title: Molecular cloning of a human gastric lipase and expression of the enzyme in yea  
A:Reference number: S07145; MUID:87299724  
A:Accession: S07145  
A:Molecule type: mRNA  
A:Residues: 1-398 <BODI>  
A:Cross-references: EMBL:X05997; NID:g31771; PIDN:CAA29413.1; PID:g758063  
A:Accession: S27102  
A:Molecule type: protein  
A:Residues: 20-24, 'X', 26, 'X', 28-33, 'X', 35-42, 'X', 44, 'X', 46-48, 'X', 50, 'X', 52, 'X', 54-56 <B  
A:Note: It is uncertain whether Met-1 or Met-7 is the initiator  
R:Bernback, S.; Blaeberg, L.  
Eur. J. Biochem. 182, 495-499, 1989  
A:Title: Human gastric lipase. The N-terminal tetrapeptide is essential for lipid bindin  
A:Reference number: S04942; MUID:89325292

A:Accession: S04942  
A:Molecule type: protein  
A:Residues: 20-45 <BER>  
C:Superfamily: triacylglycerol lipase, lingual  
C:Keywords: carboxylic ester hydrolase, extracellular protein, glycoprotein, lipid di  
F.1-19/Domains: signal sequence #status predicted <SIG>  
F.20-398/Product: triacylglycerol lipase, gastric #status experimental <MAT>  
F.34,99,185,271/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match                      49.8%; Score 1119; DB 2; Length 398;  
Best Local Similarity        52.2%; Pred. No. 7.5e-91;  
Matches    210; Conservative    73; Mismatches   107; Indels    12; Gaps     2;

QY    17 MWLLLVAMFMFORNNSVHMPKAV-----DPEAFNISITIOHGCPCEYEVEYAD 69  
        ||| : : : | : : : | : |||| : : |||| ||  
DB    1 MWLLTMASTL-----SVLGTHGLFGKLHPGSEPTVMISOMITYWGPNEEYEVWD 55  
        ||| |||| | : ||| |||| : |||| |||| |||| |||| |||| |||| ||||  
QY    70 GTILSVNRPRGLVOPPKTKGSRPVLLDHGLVGASNMYSNLPPNNSLGFIILDAGEFDWM 129  
        |||| |||| | : ||| |||| : |||| |||| |||| |||| |||| |||| ||||  
DB    56 GTILENRRLPYCKKKNSGTGRPVVELDGLASTMTLSNLPNNSLAFIILDAGIDWL 115  
        |||| |||| | : ||| |||| : |||| |||| |||| |||| |||| |||| ||||  
QY    130 GNSRGNAMSRKKTSLSDODERFAPSYDEMARFDPAYINFLIQTGOKERYVGYSGOT 189  
        ||| : : : | : |||| |||| : |||| |||| |||| |||| |||| |||| ||||  
DB    116 GMSRGNTARRRLYSPDSVERFAFSFDMATYDLPATIDIFYKKTGQQLHYVHSQGT 175  
        ||| : : : | : |||| |||| : |||| |||| |||| |||| |||| |||| ||||  
QY    190 TMGFIAFTSMPELAOKIKMYFALADPARYKHAKSPGTFELLPDMXIKGLEGGKEPELYOT 249  
        | : |||| : : |||| : : |||| : : |||| : : |||| : : |||| : : ||||  
DB    176 TIGFIASFSTNPFLAKRIKFYVALAPATVKYRKSLINKLRFPQSLEFKFIEDDKLFYPHN 235  
        | : |||| : : |||| : : |||| : : |||| : : |||| : : |||| : : ||||  
QY    250 RFLROLVTILCGOVLDDICSNIMLLGGFNPNMNMASYYAAHTLAGTSVCNIIHKRS 309  
        | : |||| : : |||| : : |||| : : |||| : : |||| : : |||| : : ||||  
DB    236 FPDOLFATEVCCREMLNLCSNALEPLICGFDSKNFTSLDVYLSHNPGTGVQNMFHT 295  
        | : |||| : : |||| : : |||| : : |||| : : |||| : : |||| : : ||||  
QY    310 QAVNSGECLAFPMGSETKLECNQTPPYRYEVRPMVTYFAMWTGODPLSPREDVKMLT 369  
        |||| : : |||| : : |||| : : |||| : : |||| : : |||| : : ||||  
DB    296 QAVKSGKFPAYDWGSEVQNRKMYDOSQPRYYVTANPPIAVNWNGSKDLLADPDQVGLL 355  
        |||| : : |||| : : |||| : : |||| : : |||| : : |||| : : ||||  
QY    370 SEVTNLIYHKNIPEMAHVDPIWGLDPAHRMYNEIHLMOOE 411  
        :: |||| : : |||| : : |||| : : |||| : : |||| : : |||| : : ||||  
DB    356 PLPLNLIYHKEIPFYNHDLFIWAMDAPQEVYNDIVSMISEDK 397  
        :: |||| : : |||| : : |||| : : |||| : : |||| : : |||| : : ||||

RESULT    6  
          733198  
          hypothetical protein ZK6.7 - Caenorhabditis elegans  
          C:Species: Caenorhabditis elegans  
          C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000  
          C:Accession: T33198  
          R:Wu, X.  
          submitted to the EMBL Data Library, May 1998  
          A:Description: The sequence of C. elegans cosmid ZK6.  
          A:Reference number: Z21301  
          A:Accession: T33198  
          A>Status: preliminary; translated from GB/EMBL/DBJ  
          A:Molecule type: DNA  
          A:Residues: 1-403 <MDX>  
          A:Cross-references: EMBL:AF067942; PIDN:AAC17694.1; GSPDB:GNO0023; CESP:ZK6.7  
          A:Experimental source: strain Bristol N2; clone ZK6  
          C:Genetics:  
          A:Gene: CESP:ZK6.7  
          A:Map position: 5  
          A:Introns: 29/3; 63/1; 219/3; 319/3; 365/3  
          C:Superfamily: triacylglycerol lipase, lingual

Query Match                      33.5%; Score 752.5; DB 2; Length 403;  
Best Local Similarity        41.1%; Pred. No. 1.7e-58;  
Matches    168; Conservative    66; Mismatches   142; Indels    33; Gaps     12;

QY    17 MM-LILIVAYMGORNVNSVHMPTKAVDPEAFNISIEITHOGYPCPEYEVADEDYITLSV 75  
        1 MMRRAVFLLAEFVQDVGVSH-----GDPELHATTPTQILRMGVPAMITYVAITDDGITLEM 55





A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-426 <NTL>  
 A:Cross-references: EMBL:281055; PIDN:CA802896.1; GSPDB:GN00022; CESP:F01G10.7  
 A:Experimental source: clone F01G10  
 C:Genetics:  
 A:Gene: CESP:F01G10.7  
 A:Map position: 4  
 A:Introns: 29/3; 64/1; 186/3; 347/2  
 C:Superfamily: triacylglycerol lipase, lingual

Query Match 29.2%; Score 655.5; DB 2; Length 426;  
 Best Local Similarity 35.9%; Pred. No. 6.6e-50;  
 Matches 140; Conservative 85; Mismatches 144; Indels 21; Gaps 9;

QY 20 LILVAVMFORNVSVNHPTKAVDEAFNISEIIIOHOGYPCPEYEVEDGYILSVNRP 79  
 DB 6 LFLILSPSINL-----SLAIDECVMTPEIGKHFGYSEVHLVPTTDEYILELHRIP 59  
 QY 80 -RGLVQPKTGSRPVYLQHGVLGASWISNLPNNSLGFLIADAGEDVMKNSGNAMS 138  
 DB 60 CKQNEKDRSSKRIYVQHGILADGFSWIPNLANSAGFVFADAGFDIWIANSRGTPAS 119  
 QY 139 RKHTLSIDODEFVAFSYDEMARFDLPVINFILOKTQEKIYVGYSGTGMFIARST 198  
 DB 120 OKHLYGEPENOKFNNFWQOSEFDLTASVDLVLETKOEFLYYGHSGQIMFSLAE 179  
 QY 199 MPBLAQIKMVFALPIATVNAKSP---STKELLIPDMIKGLFGKEF--LYQTRFL 252  
 DB 180 NPERAKIRHHALAPATVSHIGLFGFLTKFLTYAEI---LLGLPLSPISIPRTV 235  
 QY 253 RQLYVYLCGVYILDQICSNIMLLGFEFNTNNNSRASVYAHTLAGTSVONILHMSQAV 312  
 DB 236 OKMISYMSRPFMONICTLIDGIFIDG-NEKMFNOSRGVYVLCHTPAATSVKDLQHMIDLY 294  
 QY 313 NSGLRPFADMSSEFKNEKCNOPPVRRVRYDMVPTAMNGGOWMLNPEVYK-MLSE 371  
 DB 295 KSGVRSKFDYTD-GNILEYQPPPEPDELQINPTLYLWSRDDILDADTDIDRISILSK 353  
 QY 372 VTNLIYKH-NIPEWAVDFIWLDPHMY 400  
 DB 354 MNKTIAGSLERPHYSHMDVFCTHAAFDLY 383

RESULT 10  
 JT0949  
 egg-specific protein - silkworm  
 C:Species: Bombyx mori (silkworm)  
 C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 26-Feb-1998  
 C:Accession: JT0949; A28527  
 R:Sato, Y.; Yamashita, O.  
 submitted to JIPID, September 1991  
 A:Reference number: JT0949  
 A:Accession: JT0949  
 A:Molecule type: DNA  
 A:Residues: 1-559 <SAT>  
 A:Experimental source: larva  
 A:Note: this protein is a homotrimer  
 R:Indraisth, L.S.; Sasaki, T.; Yamashita, O.  
 J. Biol. Chem. 263, 1045-1051, 1988  
 A:Title: A unique protease responsible for selective degradation of a yolk protein in Bc  
 A:Reference number: A28527; MUID:88087166  
 A:Accession: A28527  
 A:Molecule type: protein  
 A:Residues: 19-40;133-144, 'T', 146-152;229-248 <IND>  
 C:Genetics:  
 A:Map position: 19  
 C:Keywords: egg yolk; homotrimer  
 F:132-133/cleavage site: Lys-Asn (egg 30.5K cysteine proteinase) #status experimental  
 F:228-229/cleavage site: Arg-Asp (egg 30.5K cysteine proteinase) #status experimental

Query Match 25.6%; Score 574.5; DB 2; Length 559;  
 Best Local Similarity 33.2%; Pred. No. 1.4e-42;  
 Matches 132; Conservative 82; Mismatches 152; Indels 31; Gaps 9;

QY 32 NSVH---MPTKAVDEAF-NMISEIIIOHOGYPCPEYEVEDGYILSVNRPGLVQPK 87  
 DB 171 NTFHDAISETORENNEDHNLNATELTKHQYVPEHFVATDGGHLYLRIP-PTHQRD 229  
 QY 88 TGSRPVYLQHGVLGASWISNLPNNSLGFLIADAGEDVMKNSGNAMSKHKTSLID 147  
 DB 230 DKRRVALLHGLGASDADWLMGSKSLAYMLSDAGDVWLVGNVKNKYSRSHSKIPA 289  
 QY 148 ODEFVAFSYDEMARFDLPVINFILOKTQEKIYVGYSGTGMFIARSTAPLAKIK 207  
 DB 290 LNDFFKFSNDELALHDLPAITIDHVLIDISQERLHIGISQATTFALMSQSPNEXIV 349  
 QY 208 MYFALPIATVNAKSP---PQTKFL-LLPDMIKGLFGKEFLYQTRFLQOLYIL 259  
 DB 350 SMHALSPIVMNVYVSPFLRMIAPTSKFYQYIHQVGHGAFEPGKHLIET-----F 400  
 QY 260 CGQVYLDQ-----ICSNIMLLGFEFNTNNNSRASVYAHTLAGTSVONILHMSQAVNS 314  
 DB 401 GGAACREKLAGRHCNNLNIVYISGINVODADIVPVMAHLPACTSARVKQYGOVNAS 460  
 QY 315 GELRAFDSSEFKNEKCNOPPVRRVRYDMVPTAMNGGOWMLNPEVYKMLSEVTN 374  
 DB 461 HPRKRYNGAET-NKRYVGASEPESYDLSKYSAPNLIHSHDAMLAHKVDKLEENLPN 519  
 QY 375 LIYHKNIPEMAH---VDFIWLDPHMYNEIHLMO 408  
 DB 520 VKOSFEVEPQGHFTDLDQFOSKKAQDYQKLMENMQ 556

RESULT 11  
 T39540  
 triglyceride lipase-cholesterol esterase - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
 C:Accession: T39540  
 C:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Beck, A.; Reinhardt, R.  
 submitted to the EMBL Data Library, February 1998  
 A:Reference number: Z21862  
 A:Accession: T39540  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-443 <WOO>  
 A:Cross-references: EMBL:AL021748; PIDN:CAAL6863.1; GSPDB:GN00067; SPDB:SPBC16A3.12C  
 A:Experimental source: strain 972h-; cosmid c16A3  
 C:Genetics:  
 A:Gene: SPDB:SPBC16A3.12C  
 A:Map position: 2  
 A:Introns: 11/2; 152/3; 282/1; 398/3  
 C:Superfamily: triacylglycerol lipase, lingual

Query Match 22.5%; Score 505; DB 2; Length 443;  
 Best Local Similarity 29.6%; Pred. No. 1.3e-36;  
 Matches 125; Conservative 91; Mismatches 154; Indels 52; Gaps 13;

QY 9 WIVSHRMEMW--LLILVAVMFORNVSVNHPTKAVDEAFNISEIIIOHOGYPCPEYEVA 66  
 DB 35 WIPRVEVMEFRVRLITVLIHYISSKTGDMTDA--OKCRNIYITCEAFGRVEHLYR 92  
 QY 67 TEDGYILSVNRPGLVQPKT-GSRPVYLQHGVLGASWIS-NLPNNSLGFLIADAG 124  
 DB 93 TQDNFILLHR-----THPKSQHREYVYCHGLMTNSELMVAVANSERSLPFLVLSG 148  
 QY 125 EDVMKNSGNAMSKHKTSLIDODEFVAFSYDEMARFDLPVINFILOKTQEKIYVGY 184  
 DB 149 YDVMIGNNGKNSKRKHITTYRKDEFPNFSLDDAMPDIDPQYDIILREGRKELNYIG 208  
 QY 165 YSOGTWMFIARSTAPLAKIKMYFALPIATVNAKSPGKFL--PDMIKGLFGK 242

```

Db 209 FSGCTAQAAMALINPDLDKNVIFGLAPATPKGFSNFVUYIKVKNKIMYH-LFGR 267
Qy 243 KEFLYQTRFLQRLQVLYICGVVLDQICSNIMLLGGFNTNNMMSRASYAATLGTATSV 302
Db 268 RCLLPVSTWQON---ICRPPIPVKIVDYSKTLFPMWDLNINSLNCKICGATLHSESSV 323
Qy 303 QNLIHNSQAVNSGELRAFDP-----WGSE-----TKNLEKCNQPTPVRYRVARDMT 346
Db 324 KSVYVHMLQIITKCTFQLYDDDMALLAGYSGRHYQVPLFPNNI-KC----- 368
Qy 347 VPTAMTGGODMLSNPDYKMLSEVTNLIYHNKINPEMAHVDIGLDAFHRYNLIHL 406
Db 369 -PULLIMGKADILINMEVMTALPPRAKEY---SIATHELDPLWGODYKKEEVPVVIDA 424
Qy 407 MQ 408
Db 425 LK 426

RESULT 12
T31611
hypothetical protein Y50E8A.g - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T31611
R:Steward, C.
submitted to the EMBL Data Library, September 1999
A:Reference number: 221047
A:Accession: T31611
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1585 <MIL>
A:Cross-References: EMBL:AL117200; NID:e1549770; PIDN:CA855050.1; CESP:Y50E8A.g
A:Experimental source: clone Y50E8A
C:Genetics:
A:Gene: CESP:Y50E8A.g
A:introns: 25/3; 60/1; 133/2; 217/3; 270/3; 337/2; 400/1; 746/2

Query Match 21.3%; Score 478.5; DB 2; Length 1585;
Best local similarity 30.6%; Pred. No. 1.8e-33;
Matches 118; Conservative 67; Mismatches 172; Indels 29; Gaps 9;

Qy 44 EAFMNISETIIHQGYPCCEYEATVEDGYILSVNRIPGLVQPKTG--SRPVLLQHGIV 101
Db 20 EETLDADTISHGYTVERKHVYTTDDGYVOLQRIPIVGRDHSILGSKRPVVFHNGLF 79
Qy 102 GGASNMISLNPNSLGLFIADAGFDVWGMNGNMSRKHKITLSDIDDEFMAFYDEMAR 161
Db 80 GSSYHFLNLPSSAAVITADAGFDVWLNIRGTETGLNHTSFSTINGVNFWNFSLEYHSH 139
Qy 162 FDLPAVINFILOKTGOEKIYVYGSGTGMGFATSTMPEL---AQIKIMYFALAPIAT 217
Db 140 YDLRQIEVLDYTRHESLFYVGHSGTA---VMFARLAEADVTWQSKIRVFALGPTAG 196
Qy 218 VHAHNSPGTKFLLPDMIMKGLFGKKEFLYQTRF-----LRQVLYICGVVLDQ 267
Db 197 FKRPMP---FLLLENYLOALI---QFALDGFGLIPVEIPRAISKPADFCSSKFTF 250
Qy 268 ICSNIMLLGGFNT--NNMMSRASYAATLGTATSVQNLHNSQAVNSGELRAFDMGSET 326
Db 251 LCSAGKVAAGIETLQGVNDSRPIILSHFSPSTSTLNLHMWQIFKYHLERLDGT-A 309
Qy 327 KNLKCNQPTPVRYRVARDMTVPTAMTGGODMLSNPDYKMLSEV--TNLIYHNKINPEM 384
Db 310 RNLIVAGKADARLEIGNIATILYFSKDDQITDEVDRREIIMQMGGLIESYLDLF 369
Qy 385 AHVDFTWGLDAPHRMNTNIIHLMQOE 410
Db 370 TFRFDITGLRATDEVKKPIYRIYKE 395

RESULT 13

```

```

T41053
triglyceride lipase-cholesterol esterase - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T41053
R:Hilbert, H.; Duesterhoeft, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, August 1998
A:Reference number: 221967
A:Accession: T41053
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-467 <HIL>
A:Cross-References: EMBL:AL031324; PIDN:CAA20447.1; GSPDB:GN00068; SPDB:SPCC1672.09
A:Experimental source: strain 972h-, cosmid c1672
C:Genetics:
A:Gene: SPDB:SPCC1672.09
A:Map position: 3
A:introns: 12/2; 164/3; 294/1

```

```

Query Match 21.1%; Score 475; DB 2; Length 467;
Best local similarity 31.2%; Pred. No. 6.4e-34;
Matches 125; Conservative 68; Mismatches 158; Indels 50; Gaps 12;

Qy 39 KAVDPEAFMNISETIIHQGYPCCEYEATVEDGYILSVNRIPR-----GLVQPKTGSRP 92
Db 73 KATHDA--DIREMKISGYVEDHLVRETDYILCIHRISKDSPGIGSPHPKTL---P 127
Qy 93 VYLQHGIVGASNMISNL--PNNSLGLFIADAGFDVWGMNGNMSRKHKITLSDIDDEF 151
Db 128 VVYCHGGLMNSVYWCNVDPRNCVLDLVNKKDYDVLGNRNKKSRLHREDSTDKEF 187
Qy 152 MAFSYDEMAREPDAVINFILOKTGOEKIYVYGSGTGMGFATSTMPELAQIKIMYFA 211
Db 188 WQFSIDDFQYDIPDTIDTILKTSQTKLYICFSQGTQAARFSLSIHPLNDKINSILA 247
Qy 212 LAPATVAKHAKSPGTFKLLPDMIMG-----LEKKEFLYQTRFLQVLYICGVV 264
Db 248 LAP-----AISPRLGNHVRVDAFVAKRPSILFELFGKRSILPSAGFWSFL----APKF 297
Qy 265 LDQICSNIMLLGGFNTNNMMSRASYAATLGTATSVQNLHNSQAVNSGELRAFDP--- 321
Db 298 FDLVLAIVCLSQLFNMSCNISYQRLVSFAHLYSYVCLVHWFIMMSAEFRMDNDQ 357
Qy 322 WGSF--TKNLEKCNQPTPVRYRVARDMTVPTAMTGGODMLSNPDYKMLSEVTNLIYK 379
Db 358 LGHDFYKLYKAKAPET-----NNIRTPILYILWGSDSL---VDIQAMNLNLPAREVH 408
Qy 380 NIPEMAHVDFWG-----LDAPHRMNTNIIHLMQOETN 413
Db 409 KVDSEHLDIMADVKKDYVPPVLRRLRDIIHHPREHEND 449

RESULT 14
E84526
probable lysosomal acid lipase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84526
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.
M.; Koo, H.; Motilal, K.S.; Cronin, L.A.; Shen, M.; Yanaiden, S.E.; Umayam, T.; Tallor
neus, D.; Nierman, W.C.; White, O.; Eilsen, J.A.; Salzberg, S.L.; Fraser, C.M.; Ventle
Nature 402, 761-768, 1999.
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: E84526
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-344 <STO>
A:Cross-References: GB:AE002093; NID:g4585908; PIDN:AAD25569.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g15230
A:Map position: 2

```

Query Match 19.8%; Score 445; DB 2; Length 344;

Best Local Similarity 31.2%; Pred. No. 1.9e-31;  
Matches 115; Conservative 55; Mismatches 124; Indels 74; Gaps 12;

```

QY 50 SEIQHOGPCPEYEATEDEGYILSYNRI---PRLGYOPKKTGSRPVYLQHLGGAS 105
   ::::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 34 ADLHPANTSCHEHSIQTDGHTLALQRYASLGR-----LQSGPVLLQHLGMAGD 86
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 106 NMISNLPNNSLGFILADAGFDVWGMNSRGNAMSRKHKTLSDODEFAFSYDEMAREFLP 165
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 87 VWFNLSPKRSISGLFILDHDFDVGWGVNRTGRYSYGHVTLSDPDKKEFWMDMSWQDLAYDLA 146
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 166 AVINFLQRTGQEKIYVYSGQTWGFIAFTSMPELQKIKMYEPALPIATVYKHAASPG 225
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 147 EMIOY- LYSINSKIFLYVHSGQTIMSFAL- TQPHVAEYVEAALLCPISTYDHTAP- 203
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 226 TFFLLPDMIMKGLFGKKEFLYQTRRLQRLVYLCGVILDOICSNIMLLGFTNNMN 285
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 204 -----LVERMV-----FMH-----LDQI----- 216
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 286 MSRAVYAHTLAGTSVONILHMSQAVNSGELRAFPMGSEPKMLEKONOPTPVRYVRDM 345
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 217 -----EYLLDYEPHPSVKNIHRLFQMIKRGTFQYDYG-YFKMLRTYGLSKPPEFILSHI 271
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 346 TVPTAMNT--GGQDWLSNEDVKMLSEVTN--LIYHKNIPEMAHVDFIIGLDAPHRY 400
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 272 PASLPWMGCGTDLADVTDEHTLAEPLSPSELY--LEDYGHIDFVLSAKEDVY 328
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 401 NEIHLMO 408
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 329 KHMIOFFR 336
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

## RESULT 15

S37965

Probable triacylglycerol lipase (EC 3.1.1.3) - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: protein YKLI40W

C:Species: *Saccharomyces cerevisiae*

C:Date: 03-May-1994 #sequence: revision 03-May-1994 #text\_change 06-Feb-1998

C:Accession: S37965; S39000; S22278

R:Namezani Rad, M.; Xu, G.; Kirchbach, L.; Filtz, C.; Keuchel, H.; Hollenberg, C.P. submitted to the Protein Sequence Database, March 1994

A:Reference number: S37953

A:Accession: S37969

A:Molecule type: DNA

A:Residues: 1-548 <RAM>

A:Cross-references: EMBL:Z28140; NID:9486238; PID:9486239; MIPS:YKLI40W

A:Experimental source: strain S288C

R:Abraham, P.R.

submitted to the EMBL Data Library, August 1993

A:Reference number: S38996

A:Accession: S39000

A:Molecule type: DNA

A:Residues: 1-548 <ABR>

A:Cross-references: EMBL:Z25464; NID:9396434; PID:9396439

R:Abraham, P.R.; Mulder, A.; van't Riet, J.; Planta, R.J.; Raue, H.A.

Yeast 8, 227-238, 1992

A:Title: Molecular cloning and physical analysis of an 8.2 kb segment of chromosome XI

A:Reference number: S22274; MUID:92245761

A:Accession: S22278

A:Molecule type: DNA

A:Residues: 1-398 <AB2>

A:Cross-references: GB:S97962

C:Genetics:

A:Gene: SGD:TGL1

A:Cross-references: SGD:S0001623; MIPS:YKLI40W

A:Map position: 111

C:Keywords: carboxylic ester hydrolase

Query Match

18.8%; Score 422.5; DB 2; Length 548;

Best Local Similarity 31.6%; Pred. No. 3.5e-29;  
Matches 119; Conservative 66; Mismatches 127; Indels 65; Gaps 15;

```

QY 61 EEEVATEDEGYILSVNRIPIGLVOP--KKTGSRPVYLQHLGGASNMISNLP- NNSIG 117
   : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 78 EDHLVATEDENYILTLHRIP-----PISKRNFNNKYVYLHGHLMCDVWCNIEHRKNLP 132
   : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 118 FILADAGFDVWGMNSRGNAMSRKHKTLSDODEFAFSYDEMAREFLPAVINFLQKTQ 177
   : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 133 FVLHDLGYDVMGMNNGNKTSTAHLNKPKNKNKFWDFSIDFAEFDPNSIEFLIDITYV 192
   : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 178 EKIVYVYSGQTWGFIAFTSMPELQKIKMYEPALPIATVK--HAK-----SPGTR 227
   : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 193 DKVYICGFSGSQQMPAASLSSEKLNRKYSHTAIPATPKGLNHRIVDTLAKSPGFM 252
   : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 228 FLLPDMIMKGLFGKKEFLYQTR--RFLRLVYLCGVILDOICSNIMLLGFTNTN 283
   : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 253 YL-----FGRKIVLPYSAVIMQRTLHPLFLENIC-----ID-IANKIL-----FNMKS 293
   : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 284 MNM-----SRASVYAHTLAGTSVONILHMSQAVNSGELRAF-----WGSEYNLEK 331
   : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 294 FNILPPOKIASYAKLYST-----TSVKSIVHMFQILRSQKFOWFESDMNLNLTTPYQI 348
   : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 332 CNOPTVRYRVDMTVPTAMTGGQDWLSNEDVKMLSEVTNLIYHKNIPEMAHVDFIM 391
   : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 349 ANPPTPT-----NIKIPILITIGIDSLVDIVAKNLP--FNSYDVAVDNIETHLDLIM 401
   : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 392 GLDAPHRYNEIHLMO 408
   : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 402 GRKDQTLVIATKVLRFIE 418
   : | : | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: January 30, 2002, 23:02:30  
Job time: 2911 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2002, 23:01:24 ; Search time 50.02 Seconds  
(without alignments)  
310.060 Million cell updates/sec

Title: US-09-578-063-47

Perfect score: 2247  
Sequence: 1 METLSRQNVSHRMEMILL.....IILMQEETNLSGCEAVL 423

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	1289	57.4	399 1	LICH_HUMAN
2	1164.5	51.8	395 1	LIPG_RAT
3	1154	51.4	398 1	LIPG_CANFA
4	1139	50.7	397 1	LIPG_BOVIN
5	1119	49.8	398 1	LIPG_HUMAN
6	663	29.5	394 1	LIP3_DROME
7	546	24.3	439 1	LIP1_DROME
8	422.5	18.8	548 1	TGIL1_YEAST
9	126	5.6	987 1	YD94_METUA
10	105	4.7	430 1	ORC2_CAEEL
11	101	4.5	436 1	PAFA_CAVPO
12	96.5	4.3	287 1	Y193_HAEIN
13	95.5	4.3	1259 1	LIN1_HUMAN
14	93.5	4.2	278 1	PRXC_STRAU
15	93	4.1	569 1	PYRD_PLAFA
16	92	4.1	748 1	PTIP_ECOLI
17	91	4.0	370 1	P2C2_SCHPO
18	90.5	4.0	1037 1	ACRD_ECOLI
19	90	4.0	277 1	BPA2_STRAU
20	90	4.0	1122 1	YG3C_YEAST
21	90	4.0	1136 1	POLG_OMV
22	89	4.0	330 1	EXOA_RHINE
23	88.5	3.9	185 1	YC2H_BACSU
24	88.5	3.9	185 1	CFAC_ECOLI
25	87	3.9	473 1	SCRB_LACIA
26	86.5	3.8	382 1	YDHH_HAEIN
27	86.5	3.8	665 1	INVA_SALTY
28	86.5	3.8	1882 1	Y468_MYCPN
29	86	3.8	525 1	ACU8_NEDCN
30	86	3.8	1866 1	YGNB_CPMV
31	85.5	3.8	428 1	Y6E3_SCHPO
32	85	3.8	429 1	ARB2_ECOLI
33	85	3.8	837 1	DBPA_DICTDI

## ALIGNMENTS

34	85	3.8	4128 1	PRKD_MOUSE
35	84.5	3.8	875 1	AMD2_XENLA
36	84	3.7	451 1	YB27_YEAST
37	84	3.7	804 1	SYL_BACSU
38	83.5	3.7	350 1	E13B_PRUPR
39	83.5	3.7	375 1	NDR3_MOUSE
40	83.5	3.7	389 1	AAT_PYRAD
41	83.5	3.7	456 1	YP95_YEAST
42	83.5	3.7	520 1	C11A_BOVIN
43	83.5	3.7	620 1	C11A_PIG
44	83.5	3.7	690 1	VGLH_HSV7J
45	83	3.7	266 1	THID_SALTY

  

RESULT 1				
LICH_HUMAN	STANDARD:	PRT:	399 AA.	
AC P38571:				
DT 01-OCT-1994 (Rel. 30, Created)				
DT 01-OCT-1994 (Rel. 30, Last sequence update)				
DT 15-JUL-1999 (Rel. 38, Last annotation update)				
DE LYSOSOMAL ACID LIPASE/CHOLESTERYL ESTER HYDROLASE PRECURSOR				
DE (EC 3.1.1.13) (IAL) (ACID CHOLESTERYL ESTER HYDROLASE) (STEROL				
DE ESTERASE) (LIPASE A) (CHOLESTERYL ESTERASE).				
GN LIPA.				
OS Homo sapiens (Human).				
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX NCBI_TaxID=9606;				
RX MEDLINE=92042192; PubMed=1718995;				
RA "Cloning and expression of cDNA encoding human lysosomal acid				
RT lipase/cholesteryl ester hydrolase. Similarities to gastric and				
RT lingual lipases."				
RL J. Biol. Chem. 266:22479-22484(1991).				
[2]				
SEQUENCE FROM N.A.				
RC TISSUE=Liver;				
RA MEDLINE=94155897; PubMed=8112342;				
RA Amels D., Merkel M., Eckerskorn C., Greten H.;				
RT "Purification, characterization and molecular cloning of human				
RT hepatic lysosomal acid lipase."				
RL Eur. J. Biochem. 219:905-914(1994).				
[3]				
VARIANT CESD/MD PRO-200.				
RA MEDLINE=94195814; PubMed=8146180;				
RA Anderson R.A., Byrum R.S., Coates P.M., Sando G.N.;				
RT "Mutations at the lysosomal acid cholesteryl ester hydrolase gene				
RT locus in Wolman disease."				
RL Proc. Natl. Acad. Sci. U.S.A. 91:2718-2722(1994).				
[4]				
VARIANTS CESD ARG-129 AND PRO-129.				
RA MEDLINE=98295576; PubMed=9633819;				
RA Ries S., Buechler C., Schindler G., Aslanidis C., Amels D., Gasche C.,				
RA Jung N., Schambach A., Fehrlinger P., Vanter M.T., Belli D.C.,				
RA Greten H., Schmitz G.;				
RT "Different missense mutations in histidine-108 of lysosomal acid				
RT lipase cause cholesteryl ester storage disease in unrelated compound				
RT heterozygous and hemizygous individuals."				
RL Hum. Mutat. 12:44-51(1998).				
-1- FUNCTION: CRUCIAL FOR THE INTRACELLULAR HYDROLYSIS OF CHOLESTERYL				
ESTERS AND TRIGLYCERIDES THAT HAVE BEEN INTERNALIZED VIA RECEPTOR-				
MEDIATED ENDOCYTOSIS OF LIPOPROTEIN PARTICLES. IMPORTANT IN				
SUPPRESSION OF HYDROXYMETHYLGUTARYL-CoA REDUCTASE AND ACTIVATION				
OF ENDOGENOUS CELLULAR CHOLESTERYL ESTER FORMATION.				
-1- CATALYTIC ACTIVITY: A STERYL ESTER + H(2)O -> A STEROL + A FATTY				
ACID.				







"The cDNA sequence encoding bovine pregastric esterase."  
 RL Gene 147,259-262(1994). TRIACYLGLYCEROL + H(2)O = DIACYLGLYCEROL +  
 CC -1- CATALYTIC ACTIVITY: TRIACYLGLYCEROL +  
 CC A FATTY ACID ANION.  
 CC -1- SIMILARITY: PARTIAL WITH OTHER LIPASES (PANCREATIC, GASTRIC,  
 CC HEPATIC, LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb-sdb.ch](mailto:license@sdb-sdb.ch)).  
 CC -----  
 CC EMBL: L26319; AAA57037.1; -  
 DR InterPro: IPR000073; Abhydrolase.  
 DR InterPro: IPR000379; Est\_lip\_thioest\_actsite.  
 DR InterPro: IPR000734; Lipase.  
 DR Pfam: PF00561; abhydrolase\_1.  
 DR PROSITE: PS00120; LIPASE\_SER\_1.  
 DR KEGG: K01120; Lipase.  
 KW Hydrolyase; Lipid degradation; Glycoprotein; signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 397  
 FT ACT\_SITE 171 171  
 FT ACT\_SITE 371 371  
 FT CARBOHYD 33 33  
 FT CARBOHYD 270 270  
 FT CARBOHYD 326 326  
 FT CARBOHYD 397 397  
 SO SEQUENCE 397 AA; 45231 MW; F68977DEB585EE36 CRC64;

Query Match 50.7%; Score 1139; DB 1; Length 397;  
 Best Local Similarity 53.7%; Pred. No. 3,7e-91;  
 Matches 212; Conservative 68; Mismatches 113; Indels 2; Gaps 2;

QY 18 WLLTVAVMFORNVNSVHPTKANDPEAFNNISRIIOHOGYPCGEYEVAEDGTLISVNR 77  
 DB 3 WLVTVCFIHNSGNACFELKRIANPASMNVSMISYWCYPSMAHYVITADGIILOVYR 62  
 QY 78 IPRGLVOPKKTGSRPVLLQHGVLGASNMVSNLNNLSLFIADAGFVWNGSRGNV 137  
 DB 63 IPHGKNNANHLGQRPVFLQHGILGSAITNMVSNLNNLSLFIADAGFVWNGSRGNV 122  
 QY 138 SRKKTISTIDODEFAWASVDEMAFDPVAVINFLQKGEKIIYVSGTGWGFIAS 197  
 DB 123 AOEHLVYSPDSPEFWAFSPDEMAEYDLPSTIDFLRTGOKLHYVHSGGTIGFIAS 182  
 QY 198 TMEPLAQIKMYFALAIATVAKHAKSPGTEFLIPDMIGLFGKKEFLVOTRFLRQ-LV 256  
 DB 183 TSPFLAEKIVFYALAAVAVKTKSLFNKLALIPHELFKIIIFGDKMF-YPHTFLEQFLG 241  
 QY 257 IYLGQVILDOICSNMILLGFGFNTNMNMSRASVYAHTLACTSVONILHMSQAVSGE 316  
 DB 242 VEMKSRTELVDLCKNALFALTGYDNKNFNKSLRDVYIAHNPATSVONTLHMQAVKSK 301  
 QY 317 LRAFDGSEFKNLEKCNQPPVRYRVDMQVTPAAMTGGODMLSNPEVKNLLSEVTNLI 376  
 DB 302 FOAFDWGAPYONLMHYHQPPPIYNLTANVPIAVVWSSADNDLADDDVDVFLSKSLNLI 361  
 QY 377 YHKNIPEMAHVDFTIGLADPHRYNETIHLMOOE 411  
 DB 362 YHKEIPRYNHIDFIWAMDAFOEYVNEIVSLMAEDK 396

RESULT 5  
 LIPG\_HUMAN STANDARD; PRT; 398 AA.  
 AC P07058;  
 DT 01-APR-1988 (rel. 07, Created)  
 DT 01-APR-1988 (rel. 07, Last sequence update)  
 DT 15-JUL-1999 (rel. 38, Last annotation update)  
 DE TRIACYLGLYCEROL LIPASE, GASTRIC PRECURSOR (EC 3.1.1.3) (GASTRIC

DE LIPASE) (GL).  
 GN LIP.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87299724; PubMed=3304425;  
 RA Bodmer M.W., Angel S., Varrault G.T., Harris T.J.R., Lyons A.,  
 RA King D.J., Pleroni G., Riviere C., Verger R., Lowe P.A.;  
 RT "Molecular cloning of a human gastric lipase and expression of the  
 RT enzyme in yeast."  
 RL Biochim. Biophys. Acta 909:237-244(1987).  
 RN [2]  
 RP SEQUENCE OF 20-45.  
 RX MEDLINE=89325292; PubMed=2753032;  
 RA Bernhaeck S., Blackberg L.;  
 RT "Human gastric lipase. The N-terminal tetrapeptide is essential for  
 RT lipid binding and lipase activity."  
 RL Eur. J. Biochem. 182:495-499(1989).  
 CC -1- CATALYTIC ACTIVITY: TRIACYLGLYCEROL + H(2)O = DIACYLGLYCEROL +  
 CC A FATTY ACID ANION.  
 CC -1- SIMILARITY: PARTIAL WITH OTHER LIPASES (PANCREATIC, GASTRIC,  
 CC HEPATIC, LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb-sdb.ch](mailto:license@sdb-sdb.ch)).  
 CC -----  
 CC EMBL: X05997; CA29413.1; -  
 DR EMBL: X05997; CA29414.1; ALF\_INIT.  
 DR EMBL: A01046; CA00125.1; -  
 DR EMBL: A12714; CA01053.1; -  
 DR PIR: S04942; S04942.  
 DR PIR: S07145; S07145.  
 DR MIM: 601980; -  
 DR InterPro: IPR000073; Abhydrolase.  
 DR InterPro: IPR000379; Est\_lip\_thioest\_actsite.  
 DR Pfam: PF00561; abhydrolase\_1.  
 DR PROSITE: PS00120; LIPASE\_SER\_1.  
 KW Hydrolyase; Lipid degradation; Glycoprotein; signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 398  
 FT ACT\_SITE 172 172  
 FT ACT\_SITE 372 372  
 FT CARBOHYD 34 34  
 FT CARBOHYD 99 99  
 FT CARBOHYD 271 271  
 FT CARBOHYD 327 327  
 SO SEQUENCE 398 AA; 45237 MW; CD3EE1621C014F0F CRC64;

Query Match 49.8%; Score 1119; DB 1; Length 398;  
 Best Local Similarity 52.2%; Pred. No. 2e-89;  
 Matches 210; Conservative 73; Mismatches 107; Indels 12; Gaps 2;

QY 17 WLLTVAVMFORNVNSVHPTKAV-----DPEAFNNISRIIOHOGYPCGEYEVAED 69  
 DB 1 WLLTVAVMFORNVNSVHPTKAV-----SVLGTHGLFGKLPSPPEVTNMSQITWGPNEEYEVTD 55  
 QY 70 GYILSVNRIPRGLVOPKKTGSRPVLLQHGVLGASNMVSNLNNLSLFIADAGFVWNG 129  
 DB 56 GYILEVNRIPRGLVOPKKTGSRPVLLQHGVLGASNMVSNLNNLSLFIADAGFVWNG 115  
 QY 130 GNSRGNASRRKKTISTIDODEFAWASVDEMAFDPVAVINFLQKGEKIIYVSGTGW 189  
 DB 116 GNSRGNASRRKKTISTIDODEFAWASVDEMAFDPVAVINFLQKGEKIIYVSGTGW 175

QY 190 TMGFIATPMPBLAOKIKMYFALPATIATYKHAASPGTKFLLPDMKIGLFGKKEFLYQT 249  
 176 TIGFIATPMPBLAOKIKMYFALPATIATYKHAASPGTKFLLPDMKIGLFGKKEFLYQT 249  
 QY 250 RFLRQVLYVLCGVIIIDQJCSNIMLLGGFNTNNMNSRASYAAHLAGTSYQVNIILHMS 309  
 236 FPDQFLATEVCSREMLNLCNSMFLTCGDSKNFTSRIDYVLSHNPAGTSVQNMHHMT 295  
 QY 310 QAVNSGELAFPMGSETKLEKCNQPTPYRYRVDMTPTPMTGQDMLSPEDYKML 369  
 296 QAVNSGELAFPMGSETKLEKCNQPTPYRYRVDMTPTPMTGQDMLSPEDYKML 369  
 QY 370 SEVTNLIYKKNIPMAHVDPIGDLAPHRNVEIILMOOE 411  
 356 PKLPNLIYKKNIPMAHVDPIGDLAPHRNVEIILMOOE 411  
 Db 356 PKLPNLIYKKNIPMAHVDPIGDLAPHRNVEIILMOOE 411

RESULT 6  
 LIP3\_DROME STANDARD; PRT; 394 AA.  
 ID LIP3\_DROME STANDARD; PRT; 394 AA.  
 AC 046108;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DE LIPASE 3 PRECURSOR (EC 3.1.1.-).  
 GN LIP3 OR CG8823.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_Taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CANTON-S;  
 RX MEDLINE=98227315; PubMed=9566193;  
 RT "The Drosophila melanogaster lipase homologs: a gene family with  
 tissue and developmental specific expression."  
 RL J. Mol. Biol. 276:877-885(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abail J.F., Adyanani A., An H.-J., Andrews-Planckoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bertan B.P., Bhandari D., Bolshakov S.,  
 RA Borova K., Botchan M.R., Bouck J., Brokstein P., Bottier P.,  
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreria S., Fleischmann W.,  
 RA Fostel C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Gloddek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasio P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merilov G., Mitsuhashi N.V., Modarres C., Morris J., Mostoslavsky A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Munz D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 CC -1- TISSUE SPECIFICITY: FAT BODY.  
 CC -1- DEVELOPMENTAL STAGE: ONLY AT LARVAL STAGES.  
 CC -1- SIMILARITY: PARTIAL WITH OTHER LIPASES (PANCREATIC, GASTRIC,  
 CC HEPATIC, LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.). ALSO SIMILAR TO  
 CC LEPIDOPTERAN EGG-SPECIFIC AND YOLK PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isdb.ch/announce/>  
 CC or send an email to [license@isdb.ch](mailto:license@isdb.ch)).

CC EMBL: Y14367; CAA74737.1; -  
 CC EMBL: AE003699; AAF54935.1; -  
 DR Flybase: FBgn0023495; LIP3.  
 DR InterPro: IPR000073; Abhydrolase.  
 DR InterPro: IPR000379; Est\_lip\_thioest\_actsite.  
 DR InterPro: IPR000734; Lipase.  
 DR Pfam: PF00561; abhydrolase; 1.  
 DR PROSITE: PS00120; LIPASE\_SER; 1.  
 KW Hydrolyase; Lipid degradation; Signal; Glycoprotein.  
 FT SIGNAL 1 20  
 FT CHAIN 21 394  
 FT ACT\_SITE 164 164  
 FT ACT\_SITE 369 369  
 FT CARBOHYD 131 131  
 SQ SEQUENCE 394 AA; 44901 MW; A718D1D743673802 CRC64;

Query Match 29.5%; Score 663; DB 1; Length 394;  
 Best Local Similarity 37.0%; Pred. No. 6; 3e-50;  
 Matches 137; Conservative 80; Mismatches 139; Indels 14; Gaps 7;

QY 51 EIIHOGYPOEYEVATEDGYIISVNRIPGLVQPK-RTGSRPVYVLOHGLVGASNNI 108  
 30 ERIEDDGYPMERREVTSDVYITLMHRIP--YSPKTESSNRPVAFMHGMLSSSDWV 86  
 QY 109 SNIPNNSLGFITLADAFDVMGSRGNANRKRKKTISTODDERMAVSIDEMAFDLPYI 168  
 87 LMPERSLAVYMLADADYDVMGNAKNTYSKAKRYPTWQIFWFSNMEIGYDVPAMI 146  
 QY 169 NFIIOTGOEKIYVYVSOOTGFIASFSTPMLAOKIKMYFALPATIATYKHAASPGTK 227  
 147 DYIATKGGQGVYVHSGOITYLVVNSRPYKIKSAHILGPAALMGNNKSLTGA 206  
 QY 228 ---FLIPDMKIGLFGKKEFLYQTRFLRQVLYVLCGVIIIDQJCSNIMLLGGFNTNN 283  
 207 FAPILQOPNAIVE-VGSMSEFMNSNFKKODIGEMCOATSPADMCANEIIFLGVDTEQ 265  
 QY 284 MNMSRASYAAHLAGTSYQVNIILHMSQAVNSGELAFPMGSETKLEKCNQPTPYRYR 343  
 266 LDVELLEHAKATSPASAVNQHFOEANSNGFRFEDY-TALRNVYEGSGFFPPYK 324  
 QY 344 DMVPTPMTGQDMLSPEDYKMLSEVYNTLIYKKNIP--EWAHVDPIGDLAPHRNVEI 401  
 325 NAKAPVLLIYGADNDMCDVSDYKALDELCPNMAALDVLVPEKNAHDLFTMGTAARKYVVD 384  
 QY 402 EIIHMOOE 411  
 385 EVLKQMSYE 394



CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces;  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92245761; PubMed=1574929;  
 RA Abraham P.R., Mulder A., Van't Riet J., Planta R.J., Raue H.A.;  
 RT "Molecular cloning and physical analysis of an 8.2 kb segment of  
 RT chromosome XI of Saccharomyces cerevisiae reveals five tightly linked  
 RT genes."  
 RL Yeast 8:227-238(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Rad M.R., Xu G., Kirchhath L., Fritz C., Keuchel H., Hollenberg C.P.;  
 RL Submitted (MAR-1994) to the EMBL/Genbank/DBJ databases.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: 225464; CAAB0958.1;  
 DR EMBL: 228140; CAAB1981.1;  
 DR PIR: S37969; S37969.  
 DR PIR: S39000; S39000.  
 DR SGD: S0001623; TGL1.  
 DR InterPro: IPR000073; Abhydrolase.  
 DR InterPro: IPR000379; Est\_1IP\_Cholest\_actsite.  
 DR Pfam: PF00561; abhydrolase.1.  
 DR HydroLase; Lipid degradation.  
 KW  
 SQ SEQUENCE 548 AA; 62979 MW; 32D1F230701CE083 CRC64;

Query Match 18.8%; Score 422.5; DB 1; Length 548;  
 Best Local Similarity 31.6%; Pred. No. 6.7e-29;  
 Matches 119; Conservative 66; Mismatches 127; Indels 65; Gaps 15;

QY 61 EEYEVATEEGYILSVNRIRGLVOP--KITGSRPVYLLDHGVLGAGSNMISNP--NNSLG 117  
 DB 78 EHLVTEEDNYLTLLRIP-----PISKRPNKKVYLLHGLMCSQVWCNTERHKNLP 132  
 QY 118 FILADAGEVYMGNSRGNAMSRKHKTLSDIDEFMAFSYDEKARFDPVAVINFILOKTCQ 177  
 DB 133 FVLHDLGYDVWGNNGNRKNTSTAHLNPKPKSKNFDFSIDFAFPDIPRSIEFIIDITKY 192  
 QY 178 EKIVYVGSQGTMGFIATSTPELAQIKMTFALAPIATVK--NAK-----SPGTR 227  
 DB 193 DKVIGISGSGAOMFAFSLSEKLNKRVSHFLAIPAMTPKGLHNRIVDTLAKSSPGFM 252  
 QY 228 FLLLDMAKGLFGKKEFLYOT---RFLRLVYILCGVYIIDQICSNIMLLGGNTNN 283  
 DB 253 YL-----FFGRKIVLPSAVIQRILHPTLFNLC---ID-IANKIL---FNMS 293  
 QY 284 NNM-----SRASYAAHTLAGTSVONILHMSQAVNSGELRAPD-----MGSETKLEK 331  
 DB 294 FHLRLRQKASTAKLST-----TSKSTIVHMFQILRSKRFQMFESDMKLSLTPPYOI 348  
 QY 332 CNOPTPRVRYRDMVTPTAMMTGODWLSNPEDVKMLLSEVTNLVILHKNIPENAHVDFTW 391  
 DB 349 ANEPFRT-----NIKIPILLIYGIDSLVDIVWKNLP--FNSVEDVAVDVNHEHLLM 401  
 QY 392 GLDAPHRKATNEIHLAQ 408  
 DB 402 GRDADTLVIAKVLRIE 418

RESULT 9  
 YD94\_METJA STANDARD: PRT: 987 AA.  
 AC 058789;

DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE HYPOTHETICAL PROTEIN MJ1394.  
 GN MJ1394.  
 OS Methanococcus jannaschii.  
 CC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;  
 CC Methanococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.V., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kellavagge A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,  
 RA Overbeek R., Kirness E.F., Weissbrock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geophagen N.S.M., Meldman J.F., Fuhrmann J.L., Nguyen D.,  
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hirst M.A., Kaine B.P., Botodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Weese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 RT jannaschii."  
 RL Science 273:1058-1073(1996).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC -1- SIMILARITY: TO M.JANNASCHII MJ1393 AND A.FULGIDUS AF2028.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U67579; AAB99404.1;  
 DR TIGR: MJ1394;  
 KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 12 32 POTENTIAL.  
 FT TRANSMEM 958 978 POTENTIAL.  
 SQ SEQUENCE 987 AA; 112360 MW; D1E628FB28CA86D CRC64;

Query Match 5.6%; Score 126; DB 1; Length 987;  
 Best Local Similarity 23.1%; Pred. No. 0.0072;  
 Matches 105; Conservative 56; Mismatches 149; Indels 144; Gaps 29;

QY 3 ETLSRQWIVSHNMKMLLVAVMFOR-----NVNSVHPTKAVDEAPMN----- 48  
 DB 340 EYDVYTKIPANKMSEVTNVLIVLKNLVPNGENVNVNIKYLNSGQFYDFNEMTGMNQ 399  
 QY 49 -----ISEIIHOG-YPCEYEVAATEDGYILSVNRIPRG--LVQPKTKGSRPVYLLQH 98  
 DB 400 YKNGIYQWSSISDHSNGNSLEKYGIST-----SLNDPNGYKLLPKREIG-RDVIY--S 450  
 QY 451 GMYVYRSNM--GGGPIDRIG--LEDENF-----GYSEVYHNSYISIDRTNGNPTE 500  
 QY 149 ---DEFMAFSYDEKARFDPVAVINFILOKTCQEKIYIVYVSGITMGFIATSTPELAK 205  
 DB 501 ISPEVYWNPNDEWYFFEL-----KI---YSNGT---ITFST----- 531  
 QY 206 IMYFALAPI-ATVYHAKSPGTRF-----LLPDMKIG---LFGKKEFLYOTR 250  
 DB 532 ---YONGSLAFTVSTIDTFTYTKFDRVYIHGCIYVYVDDLEVNSKNRDFGKANKYL-- 586  
 QY 251 FFLRLVYILCGVYIIDQICSNIMLLGG-----FNTNMMNSRASYAAHTL-----AG 299  
 DB 587 -----ETISANSEGTAVLFDDGYFKKDVNTSLN---ALNMTNITLNSNSDA 632  
 QY 300 TSVONIL-HMSQAVNSGELRAPDWSGSETKNLEKQNPVPR--YVRDMVTPTAMMTGQ 356

DB 633 TLVFNVLGNYSERDNILAKYGAFLIENYNGTNTSIRKGVASGSYSTDHGTCE 692  
 OY 357 --DWLSN---PDVYKMLSEVTNLIYHKRIPEMA 385  
 DB 693 INIWIENWTFKNDKASYFNLTNL---NI--WA 720

## RESULT 10

ORC2\_CAEEL STANDARD: PRT: 430 AA.

ID ORC2\_CAEEL

AC Q21037;

DT 01-NOV-1997 (Rel. 35, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE PROBABLE ORIGIN RECOGNITION COMPLEX SUBUNIT 2.

GN F39E10.1.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;

CC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Swaburne J.;

RL Submitted (Aug-1994) to the EMBL/Genbank/DBJ databases.

RN [2]

RP REVISIONS.

RA Jones S.J.M.;

RL Submitted (MAR-1997) to the SWISS-PROT data bank.

CC -1- FUNCTION: COMPONENT OF THE ORIGIN RECOGNITION COMPLEX (ORC) THAT BINDS ORIGINS OF REPLICATION. IT HAS A ROLE IN BOTH CHROMOSOMAL REPLICATION AND MATING TYPE TRANSCRIPTIONAL SILENCING. BINDS TO THE AS CONSENSUS SEQUENCE (ACS) OF ORIGINS OF REPLICATION IN AN ATP-DEPENDENT MANNER.

CC -1- SUBUNIT: ORC IS COMPOSED OF SIX SUBUNITS (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

CC -1- SIMILARITY: BELONGS TO THE ORC2 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to [license@sdb-sdb.ch](mailto:license@sdb-sdb.ch)).

CC EMBL: 236949; CA85415.1;

DR WormPep: F59E10.1; CE11492.

DR DNA replication: Nuclear protein.

SO SEQUENCE 430 AA; 49319 MW; 54FC086BB4AD9670 CRC64;

Query Match 4.7%; Score 105; DB 1; Length 430;

Best Local Similarity 21.9%; Pred. No. 0.15;

Matches 61; Conservative 42; Mismatches 87; Indels 88; Gaps 13;

OY 85 PAKTSSRPVVLQHLVGASNMISLDPNN-----SLGF-----TLADAGEDVV--MGN 131  
 DB 25 PEKEGRS-----OKRTGKENVASNLQSLNLEEDLEQLGFEDETVSMASAIENFMQOK 78  
 OY 132 S-----RGNAWSRKHK-----TLSTIDDEFNAFSYDEMARPDLPVAVINFILOKTOQEKY 181  
 DB 79 SASEPMNNAKSRGRRAANGNGNTEETDEEDISNALTDTKCDLPGLRNVYTKRKDKTEEEK 138  
 OY 182 YVGSQGTTFMGFIATMPELAOKIMYFALAPIATVYKHAASPGTKFLLPDMITKGLFG 241  
 DB 139 RLEHLADNDPFG-----KKKLYLAG-----FNILHGVGS 168  
 OY 242 KKEFLYQTRFLRQLYTVYLCGVVILDOICSNIMLLIGFNTN--NMNNSRASYAAHTLA 298  
 DB 169 KRDLV--TEFENELSDTYMYEMVARKDGLNVKVLGALINEMKLCNVKRG----- 217  
 OY 299 GTSVONILHWSQAV-----NSGEL-----RAFDGSE 325

DB 218 -----OSTIMARSIRKRNNSQOLLILIDNIEAPDRSD 251

## RESULT 11

PAFA\_CAVPO

ID PAFA\_CAVPO STANDARD: PRT: 436 AA.

AC P70683;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE PLATELET-ACTIVATING FACTOR ACETYLHYDROLASE PRECURSOR (EC 3.1.1.47)

DE (PAF ACETYLHYDROLASE) (PAF 2-ACYLHYDROLASE) (IDL-ASSOCIATED PHOSPHOLIPASE A2) (IDL-PLA(2)) (2-ACETYL-1-ALKYLGLYCEROPHOSPHOCHOLINE ESTERASE) (1-ALKYL-2-ACETYLGLYCEROPHOSPHOCHOLINE ESTERASE).

GN PLA2G7 OR PAFAH.

OS Cavia porcellus (Guinea pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Rodentia; Hystriognathii; Caviidae; Cavia.

OX NCBI\_TaxID=10141;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-HARTLEY; TISSUE=Liver;

RX MEDLINE=97103479; PubMed=8947850;

RA Karasawa K., Kuge O., Kawasaki K., Nishijima M., Nakano Y., Tomita M., Yokoyama K., Seta M., Nojima S.;

RT "Cloning, expression and characterization of plasma

platelet-activating factor acetylhydrolase from guinea pig.";

RL J. Biochem. 120:838-844(1996).

CC -1- FUNCTION: MODULATES THE ACTION OF PLATELET-ACTIVATING FACTOR (PAF) BY HYDROLYZING THE SN-2 ESTER BOND TO YIELD THE BIOLOGICALLY INACTIVE LYSO-PAF. HAS A SPECIFICITY FOR SUBSTRATES WITH A SHORT RESIDUE AT THE SN-2 POSITION. IT IS INACTIVE AGAINST LONG-CHAIN PHOSPHOLIPIDS.

CC -1- CATALYTIC ACTIVITY: 2-ACETYL-1-ALKYL-SN-GLYCERO-3-PHOSPHOCHOLINE + H<sub>2</sub>O -> 1-ALKYL-SN-GLYCERO-3-PHOSPHOCHOLINE + ACETATE.

CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.

CC -1- TISSUE SPECIFICITY: PLASMA.

CC -1- SIMILARITY: PARTIAL WITH OTHER LIPASES (PANCREATIC, GASTRIC, HEPATIC, LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to [license@sdb-sdb.ch](mailto:license@sdb-sdb.ch)).

CC EMBL: D67037; BA11054.1;

DR InterPro: IPR000379; Est\_lip\_thloest\_actsite.

DR InterPro: IPR000734; Lipase.

DR PROSITE: PS00120; LIPASE\_SER: 1.

DR Hydrolyase; Lipid degradation; Glycoprotein; Signal.

FT SIGNAL 1 21

FT CHAIN 22 436

FT PLATELET-ACTIVATING FACTOR

FT ACETYLHYDROLASE.

FT ACT\_SITE 271 271 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 294 294 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 349 349 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT CARBOHYD 76 76 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).

SO SEQUENCE 436 AA; 49062 MW; C359D96E392FE11 CRC64;

Query Match 4.5%; Score 101; DB 1; Length 436;

Best Local Similarity 20.4%; Pred. No. 0.33;

Matches 85; Conservative 60; Mismatches 149; Indels 122; Gaps 21;

OY 21 ILVAVYMPORNVNSVMPKAVD-PEAFNISEIIOHOGYPCPEEY-EVATED-----GX 71  
 DB 70 LMSGTTNOSFLRLRYPSQDNDPFDALM-----IPNDEYFGLTETVLGASSFLGK 119

QY 72 ILSTV-----NRIPRLVOPKRTGSR-PVYLLOHGLVGASNMISLNPNSLGFILADAGFD 126  
 DB 120 LKLLGYSVYKVPKAWNSPLKTKGKYPILIFSHG- -GAFRSI-----YSALGIELASHGFI 173  
 QY 127 V-----WMGNSRGNAWSKHKHKLTLSDIDE-----FW 152  
 DB 174 VAAVEHDESAATYYTQDAPAAESGNRSWITYYVGNLIEFERKROLQROEECSQALSW 233  
 QY 153 AFSDENARPDLPVINEFILOK-----TGOEKIYYVYGSGQTTMGFIASFTPELAQKIKM 208  
 DB 234 LLSLDEBEPKYNVLDMFNDIQGLSKDSKVALIHSHFG- -GATVIQTISE-DQFRRC 289  
 QY 209 YFAL-----APAIYKHAKSPTKFLPLDMMIKGLFGKK-ELIYQTRRLQOLV----- 256  
 DB 290 GIALDPMMFPGVEDVHSPQPLFFINSEYFQSANDTKIKKIFYPQKERRKMAIVKGSVH 349  
 QY 257 -----YILGQVILDDQICSNIMLLGGFNTNNNM-----SRASVYAHTLAGTSYQNLH 307  
 DB 350 HNFVDFTFPARGKIT-----GQMLSLKGIKIDSEVAMDLINKASLFLQKYLGLD-KNFDQ 402  
 QY 308 WSOAVNSGELRAFDMGSETKNLEKNOPTPVRYRYRDMTVPFAMWGGQDLNSPE 363  
 DB 403 WNSLME-----GDDENL-----IPEFTIPTMGSSGTGQRNPD 436  
 RESULT 12  
 Y193\_HAEIN STANDARD; PRT; 287 AA.  
 ID Y193\_HAEIN STANDARD; PRT; 287 AA.  
 AC 057427: 005013:  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PUTATIVE ESTERASE/LIPASE H10193 (EC 3.1.1.-).  
 GN H10193.  
 OS Haemophilus influenzae  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae.  
 CC Haemophilus.  
 OX NCBI\_Taxid=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RD / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 Fine L.D., Fitchman J.L., Furmann J.L., Geoghegan N.S.M.,  
 RAE Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 Venter J.C.;  
 RA "Whole-genome random sequencing and assembly of Haemophilus  
 influenzae Rd.";  
 RT Science 269:496-512(1995).  
 RL  
 CC -1- SIMILARITY: BELONGS TO A FAMILY OF ESTERASES THAT GROUPS TOGETHER  
 CC PSEUDOMONA TROPINESTERASE, DMPD, TDOF AND XYLE.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL: D32704; AAC21862.1; -  
 DR TIGR: H10193; -  
 DR InterPro: IPR000073; Abhydrolase.  
 DR InterPro: IPR000379; Est\_1ip\_chloest\_actsite.  
 DR Pfam: PF00561; abhydrolase\_1.  
 KW Hypothetical protein; Hydrolase; Serine esterase; Complete proteome.  
 FT ACT\_SITE 119 119 BY SIMILARITY.

FT ACT\_SITE 266 266 BY SIMILARITY.  
 SO SEQUENCE 287 AA; 32987 MW; F2B548619C060619 CRC64;  
 Query Match 4.3%; Score 96.5; DB 1; Length 287;  
 Best Local Similarity 19.1%; Pred. No. 0.48;  
 Matches 62; Conservative 39; Mismatches 119; Indels 105; Gaps 12;  
 QY 84 QPKTGSPPVLLDQGLVGASNMISLNPNSLGFILADAGFDVWGMGRGNANRKRKT 143  
 DB 39 QVKQPTNPVILFIHGLGDMDN-----LGLV-----ARASEHSI 75  
 QY 144 LSIODEFMAFSYDEMARFDLPAY-INFILKQTKGEKIYYVYGSGQTTMGFIASFTPEL 202  
 DB 76 LRIDRNHGHSPHSEKMYQLAEDVIAVIRHLNSKYLILGHSGGTAKKTALACPEL 135  
 QY 203 AQKIMYFALPIATVKAHKSPTKFLPLDPM-----MIKGLFGKKELYQTR 250  
 DB 136 VEK-----LVIDMSPPRYEGFGHKDVFNLFAVKNKAPEN- 171  
 QY 251 FLROLVIYLCQVILDDQICSNIMLLGGFNTNNNMRSASYAAHTLAGTSYQNLHWSQ 310  
 DB 172 --RQAKPILKQEIINDE--DVQGFMLKSFYDVSADCFEPNL---TALFNRYANIMDE- 222  
 QY 311 AVNSGELRAFDMGSETKNLEKNOPTPVRYRYRDMTVPFAMWGGQDLNSPEQKMLLS 370  
 DB 223 -----KVRVF-----TPTLFKGGNSYIKIENSEKILE 251  
 QY 371 EVTNLIYH-KNIPEMAHV---DFI 390  
 DB 252 QEPNATFTINGSGHWAIEKPDV 276  
 RESULT 13  
 LIN1\_HUMAN STANDARD; PRT; 1259 AA.  
 ID LIN1\_HUMAN STANDARD; PRT; 1259 AA.  
 AC P08547:  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 01-AUG-1988 (Rel. 08, Last annotation update)  
 DE LINE-1 REVERSE TRANSCRIPTASE-HOMOLOG.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=86230917; PubMed=2423883;  
 RA Hattori M., Kuhara S., Takenaka O., Sakaki Y.;  
 RT "LI family of repetitive DNA sequences in primates may be derived  
 RT from a sequence encoding a reverse transcriptase-related protein.";  
 RT Nature 321:625-628(1986).  
 CC -1- MISCELLANEOUS: THIS SEQUENCE WAS CONSTRUCTED FROM AN ALIGNMENT OF  
 CC PUBLISHED AND UNPUBLISHED SEQUENCES, DETERMINED IN VARIOUS  
 CC LABORATORIES, BELONGING TO THE LINE-1 FAMILY.  
 CC  
 CC PIR: A25313; GNRHL1.  
 DR InterPro: IPR000097; AP\_endonuclease\_family-1.  
 DR InterPro: IPR000300; ITPC.  
 DR InterPro: IPR000477; RYVSE.  
 DR Pfam: PF01260; AP\_endonuclease1; 1.  
 DR Pfam: PF00078; rvt; 1.  
 DR SMART: SM00128; ITPC; 1.  
 KW RNA-directed DNA polymerase.  
 SO SEQUENCE 1259 AA; 147112 MW; F3BC18A3803919F5 CRC64;  
 Query Match 4.3%; Score 95.5; DB 1; Length 1259;  
 Best Local Similarity 19.7%; Pred. No. 4.4; Indels 129; Gaps 23;  
 Matches 90; Conservative 67; Mismatches 170; Indels 129; Gaps 23;  
 QY 28 QNANVSHPMTKAVDPEAFMNI---SEIIHOGYGCCEYE----- 64  
 DB 442 QEEVESLNRPISSLEIALINSLPNKKSGPGGTAIFYQRKELVFLKLQSIENE 501

QY 65 -VATEDGYILSVNRIPR-GLVOPKKTGSRPVYL-----LOHGLVGAASNMISNL-PNN 114  
 DB 502 GILPNSFYEAISILLIPKPRDPTTKENFRISIMIDANILKLNANOIOOHKIKLHHD 561  
 QY 115 SLGFIADAGFDVWGNNSR-----GNAMSKHKTLSDIDDEFNAFSYDEM----- 159  
 DB 562 QVGFIFRANMG---WFNIRKSINIOHINTKTDNMHIIISIDEK---AFDKIQPFMLK 614  
 QY 160 -----ARPDPAVINFILOKTGOEKIYVGSQGTMGFIASFTEPDLA 203  
 DB 615 PLNKLDIGDTYLKIRAIYDKPTA-NIIL--NGQ-KLEAPLKTGRGCGPLSPILPNIV 670  
 QY 204 QIKIMFALAPATYKHAASPGKFLLLPDMAIKGLFGKKEFLYOTRFLRQLYIYICGV 263  
 DB 671 LEV-----LA--RAIRQKE-----IKGJOLKEEVEKLSLFPADMIVLENPI 711  
 QY 264 ILDOICSNIMLLGFFN-----TNMMSRASYAAHTLAGTSYONILHMSQAVNSGELR 318  
 DB 712 VSAQ---NLKLSNFSKYSKYKINQKSAFLITNNROTESOINSELPFTIASKRIKYL 768  
 QY 319 AFDWSEETKLEKCNOPTEVRYRVRDMT-----VPTAMTGGQDWLSNPDYKMLSEVT 373  
 DB 769 GIQITFDVADLFRENY-KPLANEIKEDTKMKKNIPCS-----WGRINIVKMAI--LP 818  
 QY 374 NLIYKKN-IP-----EMAHVDFIWDGADPH 397  
 DB 819 KYIYRFNAIPRIKLPMTFELETETLKFIMNQKRAH 854

## RESULT 14

PRXC\_STRAU STANDARD; PRT; 278 AA.

AC 031168;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE NON-HEME CHLOROPEROXIDASE (EC 1.11.1.10) (CHLORIDE PEROXIDASE)  
 DE (CPO-T) (CHLOROPEROXIDASE T).  
 GN CPO OR CPOF.  
 OS Streptomyces aureofaciens.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_Taxid=1894;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-TUE24;  
 RA Pelletier I., Altenbuchner J., van Pee K.-H.;  
 RL Submitted (Oct-1997) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RN X-RAY CRYSTALLOGRAPHY (1.75 ANGSTROMS).  
 RC STRAIN-TUE24;  
 RX MEDLINE=98307994; PubMed=9642069;  
 RA Hofmann B., Tolzer S., Pelletier I., Altenbuchner J., van Pee K.-H.,  
 RA Hecht H.-J.;  
 RT "Structural investigation of the cofactor-free chloroperoxidases";  
 RL J. Mol. Biol. 279:889-900(1998).  
 CC -1- CATALYTIC ACTIVITY: 2 NH + 2 CHLORIDE + H(2)O(2) -> 2 RCL + 2  
 CC H(2)O.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SIMILARITY: TO OTHER BACTERIAL NON-HEME BROMO- AND CHLORO-  
 CC PEROXIDASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; AF031242; AAB86626.1; -  
 CC PDB; 1A7U; 17-JUN-98.

DR PDB: 1A8U; 17-JUN-98.  
 DR InterPro: IPR000073; Abhydrolase.  
 DR InterPro: IPR000379; Est\_Lip\_thioest\_actsite.  
 DR Pfam: PF00561; abhydrolase; 1.  
 KW Oxidoreductase; Peroxidase; 3d-structure.  
 FT ACT\_SITE 99 BY SIMILARITY.  
 FT ACT\_SITE 229 BY SIMILARITY.  
 FT ACT\_SITE 258 BY SIMILARITY.  
 SQ SEQUENCE 278 AA; 30354 MW; 8C97A87251FBEDE5 CRC64;

Query Match 4.2%; Score 93.5; DB 1; Length 278;  
 Best Local Similarity 22.6%; Pred. No. 0.84;  
 Matches 72; Conservative 36; Mismatches 114; Indels 97; Gaps 17;

QY 91 RPVYLLOHGLVGAASNMISNLPPNSIGFLIADAGDVWNGRGNAMSKHKTLSDIDE 150  
 DB 25 QPVYLI-HQFPLSGSW-----ERQSAILDAGRGVITIDRKGFQSSQPT----- 70  
 QY 151 FNAFSYDEMAREPDLPAVINFILOKTGOEKIYVGSQGTMGFIASFTEPDLAOKIKYF 210  
 DB 71 -GYDYDFFA-----ADNTVLETLDLDAVLGVSGMT-----GEVARYVSSV- 112  
 QY 211 ALAPATYKHAASPGKFLLLPDMAIKGLFGKKEFLYOTRFLRQLYIYICGVILDOICS 270  
 DB 113 GTARIKAFAFLASL-EPELKTDDNPDGA-APKEP-----FDGIVA 151  
 QY 271 NIMLLGFFNTNMMSRASYAAHTLAGT-----SYONILHMSQAVNSGELRA----FD 321  
 DB 152 AVKADRYAFYTGFEF-----DEFYNLDENLGTIRISEAVRN--SWNTAASGGFPAALAPT 205  
 QY 322 WGESEK-NLEKCNOPTEVRYRVRDMTVPAMTGGQDWLSNPDYKMLSEVTNLIYKKN 380  
 DB 206 WYTFDRADIPRIDVPALLIHGTGDRITP-----ENTAVPFKA 244  
 QY 381 IP-EMAHVDFIWDGADPH 397  
 DB 245 LPSAEYVEVE-----GADH 258

## RESULT 15

PRXD\_PLAFA STANDARD; PRT; 569 AA.

AC Q08210;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE DIHYDROROTATE DEHYDROGENASE HOMOLOG PRECURSOR (EC 1.3.3.1)  
 DE (DIHYDROROTATE OXIDASE).  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_Taxid=5833;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-3D7;  
 RX MEDLINE=94049995; PubMed=8232427;  
 RA LeBlanc S.B., Wilson C.M.;  
 RT "The dihydroxylate dehydrogenase gene homologue of Plasmodium  
 RT falciparum";  
 RL Mol. Biochem. Parasitol. 60:349-352(1993).  
 CC -1- CATALYTIC ACTIVITY: L-DIHYDROROTATE + O(2) -> OROTATE + H(2)O(2).  
 CC -1- COFACTOR: FAD.  
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE (PROBABLE).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2002, 23:00:04 ; Search time 91.96 Seconds

(without alignments)  
672.827 Million cell updates/sec

Title: US-09-578-063-47

Perfect score: 2247  
Sequence: 1 METLSRQWIVSHREMWLL.....IHLMOQETNLSGRCEAVL 423

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL.17:\*  
1: SP\_Archaea:\*  
2: SP\_Bacteria:\*  
3: SP\_Fungi:\*  
4: SP\_Human:\*  
5: SP\_Invertebrate:\*  
6: SP\_Mammal:\*  
7: SP\_Mhc:\*  
8: SP\_Organelle:\*  
9: SP\_Phage:\*  
10: SP\_Plant:\*  
11: SP\_Rodent:\*  
12: SP\_Virus:\*  
13: SP\_Vertebrate:\*  
14: SP\_Unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1290	57.4	399	4	Q16529
2	1220.5	54.3	387	11	Q64194
3	1197.5	53.3	397	11	Q920M5
4	1134.5	50.5	395	11	Q9D798
5	1134.5	50.5	395	11	Q9D6X0
6	1134.5	50.5	395	11	Q9D6O6
7	1134.5	50.5	395	11	Q9D6P3
8	1133.5	50.4	395	11	Q9D6P8
9	1133.5	50.4	395	11	Q9D6P7
10	1131.5	50.4	395	11	Q9D6L9
11	1129.5	50.3	395	11	Q9D6T5
12	1128.5	50.2	395	11	Q9D767
13	1127.5	50.2	395	11	Q9D7C5
14	1126.5	50.1	395	11	Q9D766
15	1126.5	50.1	395	11	Q9D6L1
16	1125.5	50.1	395	11	Q9D6Q3
17	1124.5	50.0	395	11	Q9D760
18	1124.5	50.0	395	11	Q9D6S5
19	1122.5	50.0	395	11	Q9D796

20	1119.5	49.8	395	11	Q9D6N8	Q9d6n8 mus musculu
21	1116.5	49.7	395	11	Q9D6O2	Q9d6o2 mus musculu
22	1115.5	49.6	395	11	Q9D6M9	Q9d6m9 mus musculu
23	752.5	33.5	403	5	Q61866	Q61866 caenorhabd1
24	751	33.4	411	5	Q20449	Q20449 caenorhabd1
25	730	32.5	405	5	Q93789	Q93789 caenorhabd1
26	715	31.8	169	11	Q9D2L7	Q9d2l7 mus musculu
27	707.5	31.5	405	5	Q16956	Q16956 caenorhabd1
28	692	30.8	411	5	Q94252	Q94252 caenorhabd1
29	655.5	29.2	426	5	Q17766	Q17766 caenorhabd1
30	655.5	29.2	434	5	Q9VKR9	Q9vkr9 drosophila
31	639	28.4	456	5	Q9VKR5	Q9vkr5 drosophila
32	636	28.3	456	5	Q9V796	Q9v796 drosophila
33	625.5	27.8	616	5	Q77107	Q77107 plodia inte
34	588.5	26.2	838	5	Q9V8K6	Q9v8k6 drosophila
35	574.5	25.6	559	5	Q17219	Q17219 bombyx mori
36	567.5	25.3	457	5	Q9VKR2	Q9vkr2 drosophila
37	559	24.9	399	5	Q9VPE9	Q9vpe9 drosophila
38	535	23.8	416	5	Q9VQ05	Q9vq05 drosophila
39	529.5	23.6	406	5	Q9VKT7	Q9vkt7 drosophila
40	527	23.5	504	5	Q94568	Q94568 galliera me
41	505	22.5	443	3	P78898	P78898 schizosacch
42	502	22.3	435	5	Q9VCA6	Q9vca6 drosophila
43	488	21.7	355	5	Q9VKR0	Q9vkr0 drosophila
44	485	21.6	356	5	Q9VKR9	Q9vkr9 drosophila
45	478.5	21.3	401	5	Q9U276	Q9u276 caenorhabd1

## ALIGNMENTS

RESULT 1  
Q16529 PRELIMINARY; PRT; 399 AA.  
ID Q16529  
AC Q16529  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
DE LYSOSOMAL ACID LIPASE PRECURSOR.  
GN LAL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
RX NCBI\_Taxid:9606;  
RN (1)  
RP TISSUE=LIVER;  
RC SEQUENCE FROM N.A.  
RA Du H., Gregory G.A.;  
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.  
DR EMBL; Z31690; CAA83495.1; -;  
DR EMBL; U08464; AAB60328.1; -;  
DR InterPro: IPR000073; Abhydrolase.  
DR InterPro: IPR000379; Est\_lip\_thioest\_actite.  
DR InterPro: IPR000734; Lipase.  
DR Pfam: PF00561; abhydrolase; 1.  
DR PROSITE; PS00120; LIPASE\_SER; 1.  
KW Signal.  
FT CHAIN 1 21 POTENTIAL.  
FT VARIANT 22 399 LYSOSOMAL ACID LIPASE.  
FT VARIANT 16 16 T->P (IN REF. 1).  
FT VARIANT 23 23 R->G (IN REF. 1).  
SQ SEQUENCE 399 AA; 45518 MW; AE051EE098A8152B CRC64;

Query Match 57.4%; Score 1290; DB 4; Length 399;  
Best Local Similarity 61.0%; Pred. No. 2.8e-106;  
Matches 242; Conservative 59; Mismatches 94; Indels 2;

QY 15 MEMLLIVAVYMFQRNNSVMPK--AVDPEAFMNISETIIOHGYPCEEYVATEGC  
DB 1 MKMFLIVLCVILVMTLHSGSRGKLAVDPEFMNVSEIISYGFSEETIVETEDC





```

DB      61 YRIPYKKNSENIGKRPAYVLOHGLVASTNMTLPPNLSLAFILADAGYDWMGNSRGN 120
      ||| | : | ||| ||||| : ||| ||||| ||||| ||||| ||||| |||||
OY      136 AMSRKHKLISIDODEFMAFSYDEMARFDLPVAINFILOKTGOEKIYVYGSGTGMGFA 195
      ||| | : | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      121 TMSRKNVYSPDSVEFMFSEDEMAKYDLPATIDFIVOKTGEKIHVYHSGGTIGFA 180
OY      196 ESTMELAKIKMYFALAPVATVKAHAKSPGKFLLPDMKIGLFGKKEFL--YQTRFL 252
      ||| | : | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      181 FSTNPAKAKIKRFPALAPVATVKAHAKSPGKFLLPDMKIGLFGKKEFL--YQTRFL 240
OY      253 ROLVYILGGOVILDOICSNMILGFGFTNNMNSRASYAHTLACTSVONILHWSQAV 312
      ||| | : | ||| ||||| : ||| : ||| ||| ||| ||| ||| ||| |||
DB      241 GTEV---CSRELLDLGCSNALFTFCGFDKKNLNSRFDVYLGHNPAGTSTODLFHMAOLA 297
OY      313 NSGELRAFDWGSSETRKNEKCNQPTPVRYRVDMTVPATMTGGODMLSPEDVKMLLSEV 372
      ||| | : | ||| ||||| : ||| : ||| ||| ||| ||| ||| ||| |||
DB      298 KSGKLOAVYMWGSPILNMLHYNQKTPPYDVASMTVPVAVNNGHDIADPODVAALLPKL 357
OY      373 TNLVYHKNIPEMAHYDFIWMGLDAPHRMNEIHLMOQE 410
      ||| | : | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      358 PNLVYHKELIPYNHLDFIWMDADAPQEVYNEIYTMMAED 395

RESULT 6
OY      09D606 PRELIMINARY: PRT: 395 AA.
AC      09D606;
ID      01-JUN-2001 (Tremblrel. 17, Created)
DT      01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT      01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE      ADULT MALE TONGUE CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
      CLONE:2310067K20, FULL INSERT SEQUENCE.
GN      2310051B21R1K.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C57BL/6J; TISSUE=TONGUE;
      MEDLINE=21085660; PubMed=11217851;
RX      Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
      Aizawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
      Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
      Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
      Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
      Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
      Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
      Schirini L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
      Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,
      Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
      Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
      Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
      Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
      Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
      Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
      Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsuk S.,
      Hayashizaki Y.,
      "Functional annotation of a full-length mouse cDNA collection.";
RT      Nature 409:685-690(2001).
CC      -1- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DR      EMBL: AK010093; BAB26697.1; -
      MGD: MGI:191467; 2310051B21R1K.
      InterPro: IPR000073; Adhydrolase.
      InterPro: IPR000379; Est_lip_thioest_actsite.
      DR      Pfam: PF00561; abhydrolase.1.
      RA      Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsuk S.,
      Hayashizaki Y.;
      "Functional annotation of a full-length mouse cDNA collection.";
RT      Nature 409:685-690(2001).
CC      -1- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DR      EMBL: AK010120; LIPASE_SER: 1.
      RA      SEQUENCE 395 AA: 44623 MW: 2450555F5E7F707C9 CMC64;

```

```

Query Match      50.5%; Score 1134.5; DB 11; Length 395;
Best Local Similarity 53.8%; Pred. No. 1,8e-92;
Matches 214; Conservative 69; Mismatches 108; Indels 7; Gaps 3;

OY      17 MWLLVYVMO-RVNVSNHPTKAVDEAPRNSEIIOHQPEEYEVATGDIYLS 75
      ||| | : | ||| ||||| : ||| ||||| ||||| ||||| ||||| |||||
DB      1 MWLLVTVLAFAGAHGLFGKLPKPNKPNANNSQMTTYNGYSEEEVYTDGIIYG 60
OY      76 NRIPGLQVPRKTSRPPVLLQHGLVAGSANMISLPPNLSLAFILADAGYDWMGNSRGN 135
      ||| | : | ||| ||||| : ||| ||||| ||||| ||||| ||||| |||||
DB      61 YRIPYKKNSENIGKRPAYVLOHGLVASTNMTLPPNLSLAFILADAGYDWMGNSRGN 120
OY      136 AMSRKHKLISIDODEFMAFSYDEMARFDLPVAINFILOKTGOEKIYVYGSGTGMGFA 195
      ||| | : | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      121 TMSRKNVYSPDSVEFMFSEDEMAKYDLPATIDFIVOKTGEKIHVYHSGGTIGFA 180
OY      196 ESTMELAKIKMYFALAPVATVKAHAKSPGKFLLPDMKIGLFGKKEFL--YQTRFL 252
      ||| | : | ||| ||||| : ||| : ||| ||| ||| ||| ||| ||| |||
DB      298 KSGKLOAVYMWGSPILNMLHYNQKTPPYDVASMTVPVAVNNGHDIADPODVAALLPKL 357
OY      373 TNLVYHKNIPEMAHYDFIWMGLDAPHRMNEIHLMOQE 410
      ||| | : | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      358 PNLVYHKELIPYNHLDFIWMDADAPQEVYNEIYTMMAED 395

RESULT 7
OY      09D6P3 PRELIMINARY: PRT: 395 AA.
AC      09D6P3;
ID      01-JUN-2001 (Tremblrel. 17, Created)
DT      01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT      01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE      ADULT MALE TONGUE CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
      CLONE:2310069P19, FULL INSERT SEQUENCE.
GN      2310051B21R1K.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C57BL/6J; TISSUE=TONGUE;
      MEDLINE=21085660; PubMed=11217851;
RX      Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
      Aizawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
      Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
      Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
      Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
      Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
      Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
      Schirini L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
      Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,
      Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
      Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
      Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
      Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
      Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
      Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
      Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsuk S.,
      Hayashizaki Y.;
      "Functional annotation of a full-length mouse cDNA collection.";
RT      Nature 409:685-690(2001).
CC      -1- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DR      EMBL: AK010139; BAB26725.1; -

```

DR MGD;MGI:1914967;121005181RLK.  
DR InterPro:IPR000073;Abhydrolase.  
DR InterPro:IPR000379;Est\_lip\_thioest\_acetate.  
DR InterPro:IPR000734;Lipase.  
DR Pfam:PF00561;Abhydrolase\_1.  
DR PROSITE, PS00120;LIPASE\_SRR;1.  
QO SEQUENCE 335 AA; 44651 MW; E68A73CCCBFD359D CRC64

Query Match	50.5%	Score 1134.5	DB 11	Length 395
Best Local Similarity	53.8%	Pred. No. 1	8e-92	
Matches 214	Conservative 69	Mismatches 108	Indels 7	Gaps 3

[illegible]

RESULT	8.	
09C9PB8		
ID	09C9PB8	PRELIMINARY; PRT; 395 AA.
AC	09C9PB8;	
DT	01-JUN-2001 (TREMBLrel. 17, Created)	
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)	
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)	
DE	2310051B2RIK PROTEIN.	
GN	2310051B2RIK.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus	
OX	NCBI_TaxID=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	SRRAIN-C57BL/6J; TISSUE=TONGUE;	
RX	MEDLINE=21085660; PubMed=11217851;	
RA	Kawai J., Shinagawa A., Shubata K., Yoshino M., Itoh M., Ishi Y.	
RA	Alakawa T., Hara A., Fukunishi Y., Komio H., Adachi J., Fukuda S.	
RA	Alizawa K., Iizawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.	
RA	Saito T., Okazaki Y., Gobjori T., Bono H., Kasukawa T., Saito R.	
RA	Kodera K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.	
RA	Fleischmann W., Gaasterland T., Glisi C., King B., Kochia H.,	
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush	
RA	Schriell L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio	
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,	
RA	Blake J., Boffelli D., Bojunga N., Carlini P., de Bonaldo M.F.,	
RA	Brownstein M.J., Butt C., Fletcher C., Fujita M., Gariboldi M.,	
RA	Gustinchich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.	
RA	Lyons P., Marchionni L., Mashima Y., Mazzarelli J., Mombaerts P.,	
RA	Nordone P., Rilling B., Ringwald M., Rodriguez I., Sakamoto N.,	

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.  
RA Suzuki H., Toyooka K., Wang K.H., Meltz C., Whitaker C., Wilming L.,  
RA Wyushaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsuki S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection.",  
RL Nature 409:685-690(2001).  
CC -i- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.  
DR EMBL: AK010116; BAB26711.1; -.  
DR EMBL: AK009300; BAB26201.1; -.  
DR MGD: MGI:1914967; 2310051B2.R1k.  
DR InterPro: IPR000073; Abhydrolase.  
DR InterPro: IPR000379; Est\_Lip\_thioest\_actsite.  
DR InterPro: IPR000734; Lipase.  
DR Pfam: PF00561; abhydrolase; 1.  
DR PROSITE: PS00120; LIPASE\_SER; 1.  
SQ SEQUENCE 395 AA; 44579 MW; D22968655EA671D34 CR664;

Query Match	50.4%	Score 1133.5	DB 11	Length 395
Best Local Similarity	53.5%	Pred. No. 2.2e-92		
Matches 213, Conservative	70	Mismatches 108	Indels 7	Gaps 3

QY	17	MMJLLVAYMQR - RNWVSYNHPTKPAVDPEAFMNSJELLQHGQCECEYEVADEGJLTV	75
Dd	1	MMJLLVTVLSVAFGAGHGLEKGLPKPKPEAMNNSQMIITWGRSEEEYVTEDEGJLTV	60
QY	76	NRIDRGLVQPKTGRSPVLLDHLQGVGASNMJSLNLSNNSIGFLLADAGEDVWNGNSRGN	135
Dd	61	YRIFYGKNNSNIGKRRPAYLQHLGLMSATWINTLNNLSLAFLLADAGDVWNGNSRGN	120
QY	136	AMSKRHKTLSDIDDEFMAFSTDEKARDLPVAVINFLQKTQGEKITYVGSQGTMKFIA	195
Dd	121	TWSKRNYYSPEDSEVEFAFSEDEMAKXDLPATDIFVYQKTQGEKITHVGHSGGTTFIA	180
QY	196	ESTMPELQAKKKMYEFALAPITATVTHAKSPGKFLLEDMDMIKGFEGKKEFL - - - YQRRFL	255
Dd	181	ESTPMLAKKKIKRFLALPAPVATVYTESPFEKKISLIRKFLKLVTFGKNMPEPHNYLDQFL	240
QY	253	ROLVTVYLCGOVLLDIOQISNMLLGGFNTNMMNSRASVYAHAHTLAGTSVONILHMSQAV	312
Dd	241	GTEV - - - CSRLLDILLCNSMLFICGPDKKNLVNSRGVYLIGHNPACTSQDLDLPHMQLA	297
QY	313	NSGELRAFDMCSSEFKNLECKNQPLRPVRYRVKRDVTVPRAAMTGGODMLSNPDEYKMLLSEV	372
Dd	298	KSGKLGQVNMGSSPLQNNMLHYNOKRPPRYVDYSAMTVPLAVYNGGHDILADPDVAMLLPKL	357
QY	373	TNLIYHNKPEWAVHDPFWIGDAPDAHRYNNEIHLMOOE	410
Dd	358	PNLLYHKEILDYNNHLDFTWADAQOEYVNEIYVMAED	395

RESULT	9	
09C9BP7		
ID	Q9C9P7	PRELIMINARY; PRI; 395 AA.
AC	Q9C9P7;	
DT	01-JUN-2001 (TREMBLrel. 17, Created)	
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)	
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)	
DE	2310051B2IRIK PROTEIN.	
GN	2310051B2IRIK.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus	
OX	NCBI_Taxid=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C57BL/6J; TISSUE=TESTIS;	
RX	MEDLINE=21085660; PubMed=11217851;	
RA	Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,	
RA	Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,	
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamaneke I.,	
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,	
RA	Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,	

RA Fleschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schirml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Futuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Guenrich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyshew-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RA "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 CC -1- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.  
 DR EMBL: AK010236; BAB26787.1; -  
 DR EMBL: AK009413; BAB26272.1; -  
 DR EMBL: AK009428; BAB26280.1; -  
 DR EMBL: AK009459; BAB26300.1; -  
 DR EMBL: AK009473; BAB26312.1; -  
 DR EMBL: AK009474; BAB26313.1; -  
 DR EMBL: AK009479; BAB26316.1; -  
 DR EMBL: AK009523; BAB26338.1; -  
 DR EMBL: AK009525; BAB26339.1; -  
 DR EMBL: AK009546; BAB26332.1; -  
 DR EMBL: AK009571; BAB26368.1; -  
 DR EMBL: AK009573; BAB26370.1; -  
 DR EMBL: AK009729; BAB26466.1; -  
 DR EMBL: AK009773; BAB26495.1; -  
 DR EMBL: AK010019; BAB26647.1; -  
 DR EMBL: AK010035; BAB26656.1; -  
 DR EMBL: AK010058; BAB26673.1; -  
 DR EMBL: AK010061; BAB26675.1; -  
 DR EMBL: AK010124; BAB26715.1; -  
 DR EMBL: AK010125; BAB26716.1; -  
 DR MGD: MGI:191967; 2310051B21R1K.  
 DR InterPro: IPR000073; Abhydrolase.  
 DR InterPro: IPR000379; Est\_lip\_thioest\_actsite.  
 DR Pfam: PF00561; abhydrolase.1.  
 DR PROSITE: PS00120; LIPASE\_SER; 1.  
 DR SEQUENCE 395 AA; 44637 MW; D3F96B5EA671E34 CRC64;  
 SO  
 Query Match 50.4%; Score 1133.5; DB 11; Length 395;  
 Best Local Similarity 53.5%; Pred. No. 2,2e-92;  
 Matches 213; Conservative 70; Mismatches 108; Indels 7; Gaps 3;

QY 373 TNLHYKNIPMAWVDFIUGLADPHRYNEIHLMOOE 410  
 DB 358 PNLIYHKEILPEYHLDFTIWDADPOEYNEIYMAED 395  
 RESULT 10  
 ID Q9D6L9 PRELIMINARY; PRT: 395 AA.  
 AC Q9D6L9;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE ADULT MALE TONGUE CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,  
 DE CLONE:2310076L13, FULL INSERT SEQUENCE.  
 GN 2310051B21R1K.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=TONGUE;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Atakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kodota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schirml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Futuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Guenrich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyshew-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RA "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 CC -1- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.  
 DR EMBL: AK010203; BAB26766.1; -  
 DR MGD: MGI:1914967; 2310051B21R1K.  
 DR InterPro: IPR000073; Abhydrolase.  
 DR InterPro: IPR000379; Est\_lip\_thioest\_actsite.  
 DR InterPro: IPR000734; Lipase.  
 DR Pfam: PF00561; abhydrolase.1.  
 DR PROSITE: PS00120; LIPASE\_SER; 1.  
 DR SEQUENCE 395 AA; 44671 MW; E8936162510A55C CRC64;  
 SO  
 Query Match 50.4%; Score 1131.5; DB 11; Length 395;  
 Best Local Similarity 53.5%; Pred. No. 3,3e-92;  
 Matches 213; Conservative 69; Mismatches 109; Indels 7; Gaps 3;



Matches 212; Conservative 70; Mismatches 109; Indels 7; Gaps 3;

QY 17 MMLILVAVMFO-RNVNSVHMPKAVDPAPFAMNISIIIOHGPCCEYEATGDIILSV 75  
 |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :  
 Db 1 MMLILVTVSLVSAFGAHGFGKLGKPNPANNVNSOMITWGPSEYEVEDGIIILV 60

QY 76 NRIPRGLOPKKTGSRPVVLLQHGLVGASNMISNLPNNSLGFILADAGFDVWNGSRGN 135  
 |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :  
 Db 61 YRIPYKKNSENIGKRPVAVLQHGLIASAKNMTITNLPNNSLAFILADAGDVLWNGSRGN 120

QY 136 AMSRKHKTLSIDODEFAWSPDEMARFDPAYINFLQKTGGEKIYVYSGGTGFI 195  
 |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :  
 Db 121 TMSRKNVYVSPDSVEFWASPDMAKYDLPATIDFVQKTGGEKIHYVGHSGGTGFI 180

QY 196 FSTMPPLAOKIKMYFALPIATVYKAKSPGTFLPDMNIGLFGKKEFL--YOTRL 252  
 |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :  
 Db 181 FSTNPALAKKIRFYALAVATVYKTESPFKISLIPKRLKVIKFNKMFPHNYIDQPL 240

QY 253 ROLVYILGQVILDOICSNIMLLGGFNTNNMNSRASYAAHTLAGTSVONILHMSQAV 312  
 |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :  
 Db 241 GREV--CSRELLDLCSNALFFCGFDKKNLNSRFVYLGHNPAGSTQDLFHWAO 297

QY 313 NSGELRPFMGSETKNLEKNOPTPRVRYVRDMTPTAMTGGQDMLSPEDVKMLSEY 372  
 |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :  
 Db 298 KSGKLOAYWMSPLQMLNLYNOKTPPYDVSAATVPIAVWNGHDLADPODVAMLLPRL 357

QY 373 TNLITKNIPEMAHVDFTGLDAPHRMYNEIHLMOOE 410  
 |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :  
 Db 358 PNLHYKELIPYNHDLFIWAMDAPQEVYNEIYTMMAED 395

RESULT 13  
 Q9D7C5 PRELIMINARY; PRT: 395 AA.

AC Q9D7C5: 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE 2310051B21RIK PROTEIN.  
 GN 2310051B21RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=TONGUE;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 Aizawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamamoto N.,  
 Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,  
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,  
 Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 Gustinich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 Norioka P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,  
 Wyszyn-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,  
 Hayashizaki Y.;  
 RT \*Functional annotation of a full-length mouse cDNA collection.\*;  
 RL Nature 409:685-690(2001).  
 CC -i- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.  
 DR EMBL; AK009359; BAB26240.1;  
 DR MGD; MGI:1914967; 2310051B21RIK.  
 DR InterPro; IPR000073; Abhydrolase.  
 DR InterPro; IPR000379; Est\_lip\_thioest\_actsite.

DR InterPro; IPR000734; Lipase.  
 DR Pfam; PF00561; abhydrolase.1.  
 DR PROSITE; PS00120; LIPASE\_SER; 1.  
 SQ SEQUENCE 395 AA; 44748 MW; 3ED43992458DE058 CRC64;

Query Match 50.2%; Score 1127.5; DB 11; Length 395;  
 Best Local Similarity 53.3%; Pred. No. 7,4e-92;  
 Matches 212; Conservative 69; Mismatches 110; Indels 7; Gaps 3;

QY 17 MMLILVAVMFO-RNVNSVHMPKAVDPAPFAMNISIIIOHGPCCEYEATGDIILSV 75  
 |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :  
 Db 1 MMLILVTVSLVSAFGAHGFGKLGKPNPANNVNSOMITWGPSEYEVEDGIIILV 60

QY 76 NRIPRGLOPKKTGSRPVVLLQHGLVGASNMISNLPNNSLGFILADAGFDVWNGSRGN 135  
 |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :  
 Db 61 YRIPYKKNSENIGKRPVAVLQHGLIASAKNMTITNLPNNSLAFILADAGDVLWNGSRGN 120

QY 136 AMSRKHKTLSIDODEFAWSPDEMARFDPAYINFLQKTGGEKIYVYSGGTGFI 195  
 |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :  
 Db 121 TMSRKNVYVSPDSVEFWASPDMAKYDLPATIDFVQKTGGEKIHYVGHSGGTGFI 180

QY 196 FSTMPPLAOKIKMYFALPIATVYKAKSPGTFLPDMNIGLFGKKEFL--YOTRL 252  
 |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :  
 Db 181 FSTNPALAKKIRFYALAVATVYKTESPFKISLIPKRLKVIKFNKMFPHNYIDQPL 240

QY 253 ROLVYILGQVILDOICSNIMLLGGFNTNNMNSRASYAAHTLAGTSVONILHMSQAV 312  
 |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :  
 Db 241 GREV--CSRELLDLCSNALFFCGFDKKNLNSRFVYLGHNPAGSTQDLFHWAO 297

QY 313 NSGELRPFMGSETKNLEKNOPTPRVRYVRDMTPTAMTGGQDMLSPEDVKMLSEY 372  
 |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :  
 Db 298 KSGKLOAYWMSPLQMLNLYNOKTPPYDVSAATVPIAVWNGHDLADPODVAMLLPRL 357

QY 373 TNLITKNIPEMAHVDFTGLDAPHRMYNEIHLMOOE 410  
 |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :  
 Db 358 PNLHYKELIPYNHDLFIWAMDAPQEVYNEIYTMMAED 395

RESULT 14  
 Q9D766 PRELIMINARY; PRT: 395 AA.

AC Q9D766: 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE 2310051B21RIK PROTEIN.  
 GN 2310051B21RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=TONGUE;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 Aizawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamamoto N.,  
 Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,  
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,  
 Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 Gustinich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 Norioka P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,  
 Wyszyn-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,  
 Hayashizaki Y.;



RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 CC -1- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.  
 DR EMBL: AK009544; BAB26350.1;  
 DR MGI: 1914967; 2310051B21Rik.  
 DR InterPro: IPR000073; Abhydrolase.  
 DR InterPro: IPR000379; Est\_lip\_thioest\_actsite.  
 DR InterPro: IPR000734; Lipase.  
 DR Pfam: PF00561; abhydrolase\_1.  
 DR PROSITE: PS00120; LIPASE\_SER; 1.  
 DR SEQUENCE 395 AA; 44648 MW; CC69875653AA7A74 CRC64;  
 SO  
 Query Match 50.1%; Score 1126.5; DB 11; Length 395;  
 Best Local Similarity 53.3%; Pred. No. 9.1e-92;  
 Matches 212; Conservative 70; Mismatches 109; Indels 7; Gaps 3;  
 QY 17 MWLLIVAVMFO-RNVNSVHMPKAVDPEAFNISEIIOHOGYPCPEYEVATEDGYIISV 75  
 DB 1 MWLLIVTSVLSAFGGAHGLFGKLGPKNPEANNVSOMITWYMGVPSSEYEVTEDGYIIGV 60  
 QY 76 NRIPRGVQPKTKGRPVYLLQHGLVYGASNNISLPPNNSLGFILADAGFDVWNGNSRGN 135  
 DB 61 YRIPYKKNSENIGKRPVAYLLHGLIASATNITLPPNNSLAFILADAGYDWLGNRGN 120  
 QY 136 AMSRKHKLISIDQDEFMAFSYDEMAREFLPAVINFILOKGTGEKIYYGYSOGTTMGFLA 195  
 DB 121 TWSRKNVYSPDSVEFMASFSEDEMAKYDLPATIDIVYOTGGEKIHYHGSOGTTIGFLA 180  
 QY 196 FSTMELOAKIKMYFALPFIATVVKHAKSPGTELLPDMITKGLFGKKEFL--YQTRFL 252  
 DB 181 FSTNPALAKKIKRFYALPVAIVATVKTYESPKKISLIPKFLVIFGNKMFMYNYIDQFL 240  
 QY 253 ROLVYILGQVILDOIGSNIMLLGFEFNTNNMNSRASYAAHTLAGTSVQNIILMSQAV 312  
 DB 241 GREV---CSRELLDLCSNALFIKPGFKKNLNSRFVDYLLGNPAGSTODLFHMAOLA 297  
 QY 313 NSGELRAFDWGSSEKTKLEKNOPTVRYRVRDMYPTAMTGGODMLSPEDVKMLLSEV 372  
 DB 298 KSGKLOAYWMSPLONLHYNOKTPPYDYVSAMTYPIAVWNGCHDILADPOVAMLLPKL 357  
 QY 373 TNLVYHKNIPEWAVDFWIGDAPHRMYNEIITHMOE 410  
 DB 358 PNLVYHKELLPYNHLDPIWAMDAPQEVYNEIYTMMAED 395  
 RESULT 15  
 Q9DBL1 PRELIMINARY; PRT; 395 AA.  
 AC Q9DBL1.  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE ADULT MALE TONGUE CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,  
 DE CLONE:2310079020, FULL INSERT SEQUENCE.  
 GN 2310051B21Rik.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=TONGUE;  
 RX MIMLINE=210656f0; PubMed:11217851;  
 RA Kawai J., Shinaigawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,  
 RA Saito T., Okazaki Y., Gojodori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Glass C., King B., Koichi H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schirml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Donaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kaniya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilmink L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 CC -1- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.  
 DR EMBL: AK010231; BAB26784.1;  
 DR MGI: 1914967; 2310051B21Rik.  
 DR InterPro: IPR000073; Abhydrolase.  
 DR InterPro: IPR000379; Est\_lip\_thioest\_actsite.  
 DR InterPro: IPR000734; Lipase.  
 DR Pfam: PF00561; abhydrolase\_1.  
 DR PROSITE: PS00120; LIPASE\_SER; 1.  
 DR SEQUENCE 395 AA; 44659 MW; 13E0BB95FC5A40E0 CRC64;  
 SO  
 Query Match 50.1%; Score 1126.5; DB 11; Length 395;  
 Best Local Similarity 53.5%; Pred. No. 9.1e-92;  
 Matches 213; Conservative 69; Mismatches 109; Indels 7; Gaps 3;  
 QY 17 MWLLIVAVMFO-RNVNSVHMPKAVDPEAFNISEIIOHOGYPCPEYEVATEDGYIISV 75  
 DB 1 MWLLIVTSVLSAFGGAHGLFGKLGPKNPEANNVSOMITWYMGVPSSEYEVTEDGYIIGV 60  
 QY 76 NRIPRGVQPKTKGRPVYLLQHGLVYGASNNISLPPNNSLGFILADAGFDVWNGNSRGN 135  
 DB 61 YRIPYKKNSENIGKRPVAYLLHGLIASATNITLPPNNSLAFILADAGYDWLGNRGN 120  
 QY 136 AMSRKHKLISIDQDEFMAFSYDEMAREFLPAVINFILOKGTGEKIYYGYSOGTTMGFLA 195  
 DB 121 TWSRKNVYSPDSVEFMASFSEDEMAKYDLPATIDIVYOTGGEKIHYHGSOGTTIGFLA 180  
 QY 196 FSTMELOAKIKMYFALPFIATVVKHAKSPGTELLPDMITKGLFGKKEFL--YQTRFL 252  
 DB 181 FSTNPALAKKIKRFYALPVAIVATVKTYESPKKISLIPKFLVIFGNKMFMYNYIDQFL 240  
 QY 253 ROLVYILGQVILDOIGSNIMLLGFEFNTNNMNSRASYAAHTLAGTSVQNIILMSQAV 312  
 DB 241 GREV---CSRELLDLCSNALFIKPGFKKNLNSRFVDYLLGNPAGSTODLFHMAOLA 297  
 QY 313 NSGELRAFDWGSSEKTKLEKNOPTVRYRVRDMYPTAMTGGODMLSPEDVKMLLSEV 372  
 DB 298 KSGKLOAYWMSPLONLHYNOKTPPYDYVSAMTYPIAVWNGCHDILADPOVAMLLPKL 357  
 QY 373 TNLVYHKNIPEWAVDFWIGDAPHRMYNEIITHMOE 410  
 DB 358 PNLVYHKELLPYNHLDPIWAMDAPQEVYNEIYTMMAED 395

Search completed: January 30, 2002, 23:09:14  
 Job time: 550 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2002, 20:24:53 ; Search time 1858.36 seconds

(without alignments)  
11265.267 Million cell updates/sec

Title: US-09-578-063-46

Perfect score: 1269  
Sequence: 1 atgttggaacctgtctcag.....gacggtgtgagcgccgcatg 1269

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 segs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_om: \*  
20: em\_or: \*  
21: em\_ov: \*  
22: em\_pat: \*  
23: em\_ph: \*  
24: em\_pl: \*  
25: em\_ro: \*  
26: em\_sts: \*  
27: em\_sy: \*  
28: em\_un: \*  
29: em\_vi: \*  
30: em\_htgo\_hum: \*  
31: em\_htgo\_inv: \*  
32: em\_htgo\_rod: \*  
33: em\_htg\_hum: \*  
34: em\_htg\_inv: \*  
35: em\_htg\_rod: \*  
36: em\_htg\_other: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	527.8	41.6	2481	9 HSLYACLY	Z31690 H. sapiens (
2	527.8	41.6	2481	9 HSLYACLY	U08464 Human Lysos
3	527.8	41.6	2481	9 HSLYACLY	X76488 H. sapiens m
4	526.2	41.5	2493	9 HUMILPCHL	M74775 Human Lysos
5	524.6	41.3	2586	9 BC012287	BC012287 Homo sapi
6	489.2	38.6	1378	6 A26689	A26689 Precursor o
7	489.2	38.6	1378	6 A26689	A26689 Precursor o
8	470	37.0	1365	9 HSGLR	X05997 Human mRNA
9	470	37.0	1367	6 A01046	A01046 H. sapiens m
10	467.6	37.0	1367	6 A12714	A12714 Pregastric
11	467.6	36.8	1336	6 A01157	A01157 R. norvegicu
12	467.6	36.8	1335	10 RNILP	X02309 Rat mRNA fo
13	461	36.3	2358	10 MMLYACLY	Z31689 M. musculus
14	459.8	36.2	3144	10 S81497	S81497 Lysosomal a
15	458	36.1	1417	4 BOVME	L26319 Bos taurus
16	444.2	35.0	1137	6 AR039022	AR039022 Sequence
17	444.2	35.0	1137	6 AR092633	AR092633 Sequence
18	444.2	35.0	1140	6 AR039023	AR039023 Sequence
19	444.2	35.0	1140	6 AR092634	AR092634 Sequence
20	444.2	35.0	1146	6 AR039024	AR039024 Sequence
21	444.2	35.0	1146	6 AR092635	AR092635 Sequence
22	444.2	35.0	1198	6 A57760	A57760 Sequence 5
23	444.2	35.0	1528	6 A57756	A57756 Sequence 1
24	444.2	35.0	1528	6 AR039021	AR039021 Sequence
25	444.2	35.0	1528	6 AR092632	AR092632 Sequence
26	444.2	35.0	1651	4 CFLIPASE	Y13899 Canis fam11
27	441	34.8	1137	6 A39301	A39301 Sequence 2
28	441	34.8	1140	6 A39303	A39303 Sequence 4
29	441	34.8	1146	6 A39305	A39305 Sequence 6
30	441	34.8	1528	6 A39300	A39300 Sequence 1
31	388.8	30.6	1048	6 A57758	A57758 Sequence 3
32	286	22.5	699	6 AX068249	AX068249 Sequence
33	286	22.5	1308	6 AX068259	AX068259 Sequence
34	269	21.2	43456	2 AL353113	AL353113 Human DNA
35	269	21.2	76653	2 AC015506	AC015506 Homo sapi
36	269	21.2	76653	2 AC015506	AC015506 Homo sapi
37	204.4	16.1	148720	2 AC019353	AC019353 Homo sapi
38	204.4	16.1	172834	2 AC011277	AC011277 Homo sapi
39	151.2	11.9	76653	2 AC015506	AC015506 Homo sapi
40	147	11.6	161826	9 AL358532	AL358532 Human DNA
41	140.6	11.1	401	9 HSLIPCHL04	U04288 Human Lysos
42	140.6	11.1	1851	9 HSLIP4	X75491 H. sapiens L
43	140.6	11.1	169404	2 AL353751	AL353751 Homo sapi
44	140.6	11.1	188147	2 AL353533	AL353533 Homo sapi
45	127.6	10.1	40114	3 CELZK6	AF067942 Caenorhab

## ALIGNMENTS

RESULT	1	
LOCUS	HSLYACLY	2481 bp mRNA
DEFINITION	H. sapiens (HepG2) LAL mRNA for lysosomal acid lipase.	17-FEB-1997
ACCESSION	Z31690	
VERSION	Z31690.1	GI:506430
KEYWORDS	LAL; Lipase; Lysosomal acid lipase.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	Du, H. and Gregory, G.A.	
TITLE	Structural Conservation of Putative Functional Motifs between Mouse and Human Lysosomal Acid Lipase	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 2481)	
AUTHORS	Du, H.	
TITLE	Direct Submission	
JOURNAL	Submitted (05-APR-1994) Hong Du, Division of Human Genetics,	



REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1 (bases 1 to 2481)  
Du, H. and Gregory, G.A.  
Structural Conservation of Putative Functional Motifs between Mouse and Human Lysosomal Acid Lipase  
Unpublished  
2 (bases 1 to 2481)  
Du, H.  
Direct Submission  
Submitted (08-APR-1994) Hong Du, Division of Human Genetics,  
Children's Hospital Medical Center, 3333 Bernet Street, Cincinnati,  
OH 45229-3039, USA

FEATURES  
source  
Location/Qualifiers

1. .2481  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="clone hLAL"  
/clone\_1lb="HepG2 cDNA library/Lambda ZAP"  
/cell\_line="HepG2"  
/cell\_type="hepatoma"  
/tissue\_type="liver"  
1. .33  
34. .96  
34. .1233  
/codon\_start=-1  
/product="lysosomal acid lipase"  
/protein\_id="A860328.1"  
/db\_xref="GI:505053"  
/translation="MKRPLGLVYCLVLTLSHSGSKGKITANDPENNAVSEIISW  
GPSEYLVETEDGYILCLNRIPIGRKNSDKPKPVFLQHLGLADSSMWNILANS  
SLGFLIADGFDWNGNSRGNTSRKHKILSVSODEFWASYSDEMAKYDLPASINFL  
NKTGQOYVYVGHSGCTTGFIAPSOIPELAKRIKMFALGPVASYAFCTSPMAKLR  
LPHLIKDLGDEKFLPOSALFKMLGTHVCTHVLKELGKELGFLGFMERNLMSR  
VDYTTTHSPAGTSVOMLHMSOAVKFOKPOAPMGSSAKKVPFYHNSYPTVYVKDL  
VPRVMSGHMDLADYDVNLTITLTNLTNLFHESIPWEHLDPITWGLDAPWRLNKKIT  
NMRRKYQ"

5'UTR  
sig-peptide  
CDS  
variation  
mat-peptide  
variation  
3'UTR  
polyA\_signal  
polyA\_site  
BASE COUNT  
ORIGIN

692 a 477 c 513 g 799 t

Query Match 41.6%; Score 527.8; DB 9; Length 2481;  
Best Local Similarity 65.4%; Pred. No. 7.3e-147;  
Matches 791; Conservative 0; Mismatches 412; Indels 6; Gaps 1;

QY 39 cagaatggaatgtgcttctgattctgtgtggtatatactccagagaatgtgaattc 98  
DB 30 CAGAAATGAATAATGCGGTTCTTGCGGTGCGTCTGTTGTTCTCGACCCCTGCAATTC 89  
QY 99 ag-----taccatgccaactaaagctgtgaccagagaagcatcctgaatatatga 152  
DB 90 TGAGGGGTCTAGAGGGAATCTGACAGCTGTGAGATCTGANAACAACATGAAATGTGAGGA 149  
QY 153 aatcacaacatcaagaagctatccctgtgaggaatataatgaatcgaactgaagaatgga 212  
DB 150 AATTATCTCTTACTGGGATTCCTCTAGTGAAGATACCTGATTGAGACAGAGATGATA 209  
QY 213 tatccttctgttaacagaattccctgagacctagtcgaactaagaagacagttccag 272  
DB 210 TATTCTGTGCTTACCGAATTCCTCATGAGAGAGAAACCATTCGACAAAGATGCCAA 269  
QY 273 gacctgtgttactgacagatgacctgtgaggtgaggtgacgaactgaattccaact 332  
DB 270 ACCAGTTGCTTCTCGAACAATGCTGTCGAGGATTCATGAACTGCGGTGCACAAACCT 329

QY 333 gccacaataagcctgggtcttactcttgcagagatgctgtttgacgtgtgagtgaggaa 392  
DB 330 TGCCAAAGACAGCGCTGGCTTCATCTCTGTGATGATGCTGTTTGACGTGTGATGGGCAA 389  
QY 393 cagcagagggaacgcctggtctcgaataaacacagaacactctccatagaaccaatgagt 452  
DB 390 CAGCAGAGAAATACCTGCTCGTCCGAAACATTAACACTCTCAGTTTCTCAGATGAAAT 449  
QY 453 ctgggtcttgaattatgattgattgattgattgattgattgattgattgattgattgatt 512  
DB 450 CTGGCTTTCAGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 509  
QY 513 ttgcagaacaaagcggccaggaagaatcattatgctgcgtatccacagggcaccacact 572  
DB 510 TCTGAATTAACCTGGCCAAAGAAAGTATATATGATGATGATGATGATGATGATGATGAT 569  
QY 573 gggcttattgcatcttccaccatgcccagagctggtcagaagaatcaaatgattcttc 632  
DB 570 AGGTTTATATACATTTTCCACAGATCCCTGAGCTGCTAAAGGATTTAAATGTTTTC 629  
QY 633 tttagcaccatagcactgtttaagcatagcaaaaagcccgaggacaaattttgtgct 692  
DB 630 CTTGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 689  
QY 693 gccagatatgatacaaggagattgttgcgaataaagaatcttgatcacagaccagatt 752  
DB 690 ACCAGATCATCTATTAAAGACTTTTGAAGACTTTTGAAGACTTTTGAAGACTTTTGAAG 749  
QY 753 tctcagaacactgttattacacttctgtgcaggtgattcttgcagatttgaagtaa 812  
DB 750 TTTAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 809  
QY 813 tatctgttacttctgtgtgatttcaacacacacacacacacacacacacacacacacac 872  
DB 810 TCTGTTGTTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 869  
QY 873 atatgctgccacactcttctgtgaacatctgtgcgaataatacttcaactgtgaagcaggc 932  
DB 870 ATATACAAACATCTCTCTGCTGAGACTTCTGTCAAAACATGTTACATCTGAGACCGGC 929  
QY 933 agtgaattctgtgaactcgggcatcttgcattgagtgagtgagtgagtgagtgagtgag 992  
DB 930 TGTAAATTTCCAAAGTTTCAAGCTTTGAGCTGGGAGAGCTGCCAAGAAATTTTTC 989  
QY 993 atgcaatcagcgaactctgtgaaggtacagagtcagagatgacggtccctacagcaat 1052  
DB 990 TTACACACAGATTATCTCTCCACATACATGATGATGATGATGATGATGATGATGAT 1049  
QY 1053 gtgacagagagtgacagactgttccaaatccagaagaagctggaataatgtctctctga 1112  
DB 1050 CTGAGAGGGGGGTGACAGCTGCTGTGCAAGATGCTAGAGACTCAATTTCTAGAGCTCA 1109  
QY 1113 ggtgaccaactcattcaccataagaatatctcgaatgggttcaagtgatttcaatctg 1172  
DB 1110 GATCACCAACTGTTGTTTCCATGATGAGACATTCGGAATGGAGCATCTTGACTCATTTG 1169  
QY 1173 ggggttggatgctcaccacgtatgtaaatgaatcattcatttgaatgagcagggagga 1232  
DB 1170 GGGCTCTGATGCTCCCTTGAGGCTTTTAAATTAATTAATTAATTAATTAATTAAT 1229  
QY 1233 gaccacact 1241  
DB 1230 GTGAAGCT 1238

RESULT 3  
LOCUS HSLAL 2626 bp mRNA PRI 25-FEB-1994  
DEFINITION H.sapiens mRNA for lysosomal acid lipase.  
ACCESSION X76488.1 GI:434305  
VERSION  
KEYWORDS lysosomal acid lipase.

SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 2626)  
AUTHORS Ameis, D., Merkel, M., Eckerskorn, C. and Grottel, H.  
TITLE Purification, characterization and molecular cloning of human hepatic lysosomal acid lipase  
JOURNAL Eur. J. Biochem. 219 (3), 905-914 (1994)  
MEDLINE 94155897  
REFERENCE 2 (bases 1 to 2626)  
AUTHORS Ameis, D.  
TITLE Direct Submission  
JOURNAL Submitted (29-NOV-1993) D. Ameis, Medical Department, University Hospital Lippendorf, Martinistrasse 52, 20246 Hamburg, FRG

FEATURES  
source  
1. 2626  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/tissue\_type="liver"  
146. 1345  
/gene="LAL/LIPA"  
146. 1345  
/gene="LAL/LIPA"  
/EC\_number="3.1.1.13"  
/note="sterol esterase"  
/codon\_start=1  
/evidence="experimental"  
/product="lysosomal acid lipase"  
/protein\_id="CA54026.1"  
/db\_xref="GI:434306"  
/db\_xref="SWISS-PROT:P38571"  
/translational="MKRFLGLVCLVLTMLHSESGKLTAVDPETNNVSEIISYM  
GPESEYLVETEDGYILCLRIPIHGRKNSDKPKPYVELOHGLDSSNMVNLNS  
SLGFLIADGPDVWGNRSNTSRKHTLSDQDEPMFASVDEMAKXDLPAISINFL  
NKTGCOQVYVGHSGCTTGTFIAFSOIPLEAKRIKFPALGPVASYAFCTSPAKRGR  
LPDHLTKDLFGDERLPQSAFLKMLGTHCTHYILAEIGNLCLFLCGFNERLNLSR  
VDVYTHSPAGTSVQNMHLMSQAVKFOKQAFDWGSSANRYHNSYPTPVNKKML  
VPLAVWSGGHMDLADYVDVNIILTOITNLFHESIPWEHLDFIMGLDAPWRLNKKII  
NLNRKYQ"  
146. 226  
/gene="LAL/LIPA"  
/evidence="experimental"  
227. 373  
/gene="LAL/LIPA"  
/note="propeptide"  
/evidence="experimental"  
374. 1342  
/gene="LAL/LIPA"  
/evidence="experimental"  
/product="lysosomal acid lipase"

BASE COUNT 741 a 527 c 547 g 811 t  
ORIGIN

Query Match 41.6%; Score 527.8; DB 9; Length 2626;  
Best Local Similarity 65.4%; Pred. No. 7.4e-147;  
Matches 791; Conservative 0; Mismatches 412; Indels 6; Gaps 1;

Db 322 TATTCTGCGCTTAACCGAATTCCTCATGGAGAGAAACAACTTGACAAAGTCCCAA 381  
Oy 273 gacctggtgttctctgcagcaatgacctagtgtgaagtgctcagaactggtttccaact 332  
Db 382 ACCAGTGTCTTCTCTGCAACATGCTGCTGCGAGATTCAGTAACAGGCTCCAAACCT 441  
Oy 333 gcccaacaatagctgtgctcattctcgtcagatgctggtgttgagctgtgagtgagaa 392  
Db 442 TGCCACACGACGCTGGCGTTCATTCCTGCTGATGCTGTTGACCTGTGATGGGCA 501  
Oy 393 cagcaggggaaacgcctgctcgaacaaacagacacatccatagaccacaagaatgagt 452  
Db 502 CACGACAGGAATACCTGGCTGCGAACAATAGACACTCTCAGTTCTCAGATGAAAT 561  
Oy 453 ctggcttctcagttatgatatgatatgctgagtttgacctctcgcagtgataacttatt 512  
Db 562 CTGGGCTTTCAGTTATGATGATGAGCAAAATATGACCTACCAAGCTTCCTTAACCTCAT 621  
Oy 513 ttgcagaaaacgycgcagaaaagatctattatgtcgtctattcaagggcaccacat 572  
Db 622 TCTGAATAAACCTGCCAAGACAGTGTATTATGTGGGTCAATTCCTAAAGCACACTAT 681  
Oy 573 gggcttattgcatcttccacatgcccagagctcgcagaaaatcaaatgattttgc 632  
Db 682 AGGTTTATAGCATTTTTCACAGATCCCTGAGCTGCTAAAGATTAATGTTTTTTC 741  
Oy 633 tttagcaccatagaccactgtttaagcatgcaaaagcccgaggacaaattttgtgtc 692  
Db 742 CCGGGGTCCGTGGCTTCGCTCGCTCTGTACTAGCCCTTAAGCCAAATTAGACAGATT 801  
Oy 693 gccagatatagtatcaaggagattgttgcaaaaagaatttcgtatacagaccagatt 752  
Db 802 ACCAGATCATCTATTATGAGCACTATTATGAGACAAAGAAATTTCTCCAGAGTGGCT 861  
Oy 753 tctcaagaacttgattttaacctgtgagcaggtgattctcgtatcagatttgagtaa 812  
Db 862 TTGAAAGTGGCTGGTACCCACGCTTGACCTCATGTATCTGAAAGAGCTCTGTGAAA 921  
Oy 813 taccatgttactctcgtggtgattccaacacaaacataatgaacatgagccagcaagt 872  
Db 922 TCTCTGTTTCTCTGCTGTGATTAATGAGAAATTTAAATGTCTAGAGTGGATGT 981  
Oy 873 atatgctgcccacactctctgtggaacatctgtgcaaaatattctaacctggagccagc 932  
Db 982 ATATACACACATCTCTGCTGTAACCTGTGCAAAACATGTTACACTGGACCCAGGC 1041  
Oy 933 agtgaattcgtggaactccggcattgactgagggagtgagagccaaactcgaanaa 992  
Db 1042 TGTTAATTCCAAAAGTTTAAAGCCTTTGACTGGGGAAGCAGTGCCAAAGAAATTTTCA 1101  
Oy 993 atgcaatcagccacactctctgtaagtgatcaagatcagaatatacaggttccctacaagaat 1052  
Db 1102 TTACAAACAGATTAATCTCCACATACATGTAAGAGACATGCTGTGCGACATCGAGT 1161  
Oy 1053 gtgagcagaggtgtcagagctggtttcaatccagaagagctgtaaaatgtgtctctga 1112  
Db 1162 CTGAGAGGGGGGTACACATGCTGTCGATGCTACAGAGTCAAAATATCTTACTGACTCA 1221  
Oy 1113 ggtgaccaactcatctacataagaatatctcgtgaattggtgctcaactgattcactctg 1172  
Db 1222 GATCACCAATTTGGTGTTCATGACAGACATTCGGGAATGGAGACATTTACTCTCATTTG 1281  
Oy 1173 ggggttgatgtctctccacggtatgtcaaatgaatcaatccatctgtatcagcagagga 1232  
Db 1282 GGGCTGTGATGCCCTTGAGAGCTTTATTAATAATTAATTAATTAATTAATTAATTA 1341  
Oy 1233 gaccacact 1241  
Db 1342 GTGAAGACT 1350

RESULT 4



QY	1173	gggcttggagcctccaccgtagtcatgaatacatcatcattgatgcagcagagga	1232
Db	1177	gggcttggagcctccaccgtagtcatgaatacatcatcattgatgcagcagagga	1236
QY	1233	gaccacct	1241
Db	1237	GTGAAGCT	1245
RESULT	5		
LOCUS	BC012287		
DEFINITION	BC012287	2586 bp mRNA	PRI 08-AUG-2001
ACCESSION	BC012287	Human sapiens, similar to lipase A, lysosomal acid, cholesterol esterase (Wolman disease), clone MGC:5229 IMAGE:2900168, mRNA, complete cds.	
VERSION	BC012287.1	GI:15126726	
KEYWORDS		MGC.	
SOURCE		human.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS		1 (bases 1 to 2586)	
TITLE		Strausberg, R.	
JOURNAL		Direct Submission	
REMARK		Submitted (06-AUG-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	
COMMENT		NIH-MGC Project URL: <a href="http://mgc.ncl.nih.gov">http://mgc.ncl.nih.gov</a> Contact: MGC help desk Email: <a href="mailto:cgabbs-remail.nih.gov">cgabbs-remail.nih.gov</a> Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. CDNA library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: <a href="http://www.hgsc.bcm.tmc.edu/cdna/">http://www.hgsc.bcm.tmc.edu/cdna/</a> Contact: Villalobos, L.M., Luna, R.A., Hale, S.M., Hui, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.	
FEATURES		Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <a href="http://image.lnl.gov">http://image.lnl.gov</a> Series: IRAX Plate: 3 Row: c Column: 3 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 434305. Location/Qualifiers 1..2586 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="MGC:5229 IMAGE:2900168" /tissue_type="Placenta, choriochorionoma" /clone_id="N1H_MGC_10" /lab_host="DH10B" /note="Vector: pCMV-SPORT6" 101..1300 /codon_start=1 /product="Similar to lipase A, lysosomal acid, cholesterol esterase (Wolman disease)" /protein_id="AAH12287.1" /db_xref="GI:15126727" /translation="MKMRPLGIVYCLVIMPLHSRSGSKLTALDPETRMNVSEITIRVGFPEEPLVETEDGYIICLNRIPIGRNHSKDGKPVVFLQHGILLAOSSNVTNVLANSSTGLTADAGFDVWNGSRNTWSRKRTLSVQDEWAFSEIDEMAKYDIPASINFIILNKGQEQEVYVYHSGQTTIGFTAFSAQIPELAKRIKMFALGPVSAVFCETSPMKLGLPRLHIDLFDEKDFELPOSALFKVGHGCHVHILELTCGMLFCILGSPRNINMSVDVYTHSPAGTSQONLHMSQAFKPKPOAFDMGSSAKNYFHNOSYPPRYNKKDMLVPAAVMSGCHMDLADYDVNIIILQITNLVFIHSEIPEHEHIDFTWIGDAPWRLTYNKIILNMRKYYG"	
CDS			
BASE COUNT	736 a	531 g	808 t

Query Match	41.3%	Score 524.6	DB 9	Length 2586
Best Local Similarity	65.3%	Pred. No. 6,7e-146		
Matches 789	Conservative	0	Mismatches 414	Indels 6
				Gaps 1
ORIGIN				
Query	39	cagaatgggaatgtggtcttcgtatcttggtggtgtatgtcttcagagaatgtgaatc	98	
DB	97	CAGAAATGAATAATCCGGCTTCTTGGGGTGGGGTGGCTCTTTGGCTTCCTGGCCCCCTGCATTC	156	
QY	99	ag-----taccatgtccaaactaaagtctgtgagcccaagaagcatcattatattgta	152	
DB	157	TGAGGGGCTCTGGAGGGAACCTGACACCTTTGGATCCTGGAACAAATGAAATGTGTGATA	216	
QY	153	aatcalccacaactcaaggctatcccctgttgaaggataatgaatgcgaactgaaatgtgta	212	
DB	217	AATTAATCTCTTAAGTGGGGATTTCCCTATGTGAGGAATCCATAGTTGAGACGAAGATGTGATA	276	
QY	213	tatcccttcgtttaaagaagatccctgcagggcctagtgtgaacctaaagaacagttccag	272	
DB	277	TATCTGTGCTCTTAACCGAATTCCTATGTGGAGGAAGAACCAATTCGACAAAGATCCCA	336	
QY	273	gacctgtggttaccatcaacatgacctagtgtgaagttgtaaacatgtattccaact	332	
DB	337	ACCAAGTGTCTTCTCCGCAACATGGCTTCTGCGCAGATTTCTACTAATCGGGTACAAACT	396	
QY	333	gcccaacaataagcttgggtcttcatcttggcagaatgctgttttttgacgtgtgagtgggaa	392	
DB	397	TGCAACACAGCAGCTGGGGCTTCTATCTTCTGCTATGCTGTGTTGAGGTGTGATGGCA	456	
QY	393	cagcagagggaaacgcctgtgtctcgaaacacagaacactctccatagacaaagtga	452	
DB	457	CAGCAGAGGAAATTAACCTGTCTCGGAAACATTAAGACACTCTCAGTTTCTCAGAGTAAT	516	
QY	453	ctgggcttccagtatgaatgagatggcgaagtgttaccttccctgcgaagtataacttat	512	
DB	517	CTGGGCTTCACTTATGATGAGATGGCAAAATATACCTCAGCTTCCATTAACTTAT	576	
QY	513	tttgcagaanaaaggcggaagaaagatctattatgtcgctatctacagagcaccacat	572	
DB	577	TCTGAATAAACATGGCCACAAGAACAGTGTATATGTGGTCAATTCACAGCACACAT	636	
QY	573	gggcttattgtcatcttccacatgtccagagctgtgctcagaanaatcaaatgtattgtgc	632	
DB	637	AGGTTTATAGCATTTTACAGAGTCCCTGAGCTGCTAAAGAGATTAAATGTTTTTTC	696	
QY	633	ttttagcaccatagcactgttaagcatgtgcaaaaagccccgggaaccaattttgtgtc	692	
DB	697	CCTGGGCTCTGGGCTTCCGCTCGCTTGTATAGCCCTATAGGCAATTAAGACGAT	756	
QY	693	gccagataatgtatcaagagatgttctgtgcaaaaagaattctgtatcagacaga	752	
DB	757	ACCAATATCATCATTAAGAGACTTATTTTGGAGACAAATAATTTCTTCCACAGATGGCT	816	
QY	753	tctcagacaactgttatttaaccttgtgtgcaggtgtatcttgatcagattgtgata	812	
DB	817	TTTGAAGTGGCTGGGTACCCACGTTTGACATCAATGTACTGAAGGAGCTGTGGAA	876	
QY	813	tatcatgttactctgggtgattccaacaacaacatctgaacatgtgacgaagtgt	872	
DB	877	TCTCTGTTTCTTCTGTGTGATTTAAAGAGAAATTTAAATATGTCTAAGTGGAGT	936	
QY	873	ataatgtgcccaactctgtctggaacatctgttcaaatatcttcaactgtgagcca	932	
DB	937	ATATACAAACATTTCTCTGCTGTGAATCTGTGTCAAAACATGTTTACACTGTGAGCCAGC	996	
QY	933	agtgaatctgtgtgaactcgggcatcttgaactgtgggagtgatgagccaanaatctgaaa	992	
DB	997	TGTTAAATCCAAAGTTTCAAGCTTGTACTGGGGAAGCGTCCCAAGATTTATTTCA	1056	
QY	993	atggaatcagccaactctgtgaagtacagatcagaatataatgcgtcccttaagaat	1052	



```

Db 1057 TTACAACAGAGATATCTCCACATACATGTAAGAGACATGCTTGCCAGCTGACGT 1116
Qy 1053 gtgacagaggagtcagagcttgcttcaatccaaagaacggaatgctctctcga 1112
Db 1117 CTGGAGCGGGGTCACGACTGGCTTGACAGATGCTACAGACGCAATATCTTACTACTCA 1176
Qy 1113 ggtgacacacatcctcctccatgaagaatctctgaagggagctgagttactctcgt 1172
Db 1177 GATCACCACTTGCTGCTTCATGAGAGATTCGGAAATGGAGCATCTTGACTTCATTG 1236
Qy 1173 gggttgagtgctcctccacgtaigtacaaatgaaatccatctgagtcagcagaaga 1232
Db 1237 GGGCGTGATGGCCCTTGAGAGCTTTATAATAATTAATTAATTAATGAGAAATATCA 1296
Qy 1233 gaccacact 1241
Db 1297 GTGAAGCT 1305

RESULT 6
LOCUS A26689 1378 bp DNA PAT 05-APR-1995
DEFINITION Precursor of rabbit gastric lipase coding sequence.
ACCESSION A26689
VERSION A26689.1 GI:905029
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct.
artificial sequence.
REFERENCE 1 (bases 1 to 1378)
AUTHORS Benicourt,C., Blanchard,C. and Junien,J.L.
TITLE Recombinant gastric lipase from rabbit and pharmaceutical
compositions
JOURNAL Patent: EP 0542629-A 9 19-MAY-1993;
INSTITUT DE RECHERCHE JOUVEINALE
LOCATION/Qualifiers
FEATURES
SOURCE
1..1378
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT 397 a 287 c 276 g 418 t
ORIGIN
Query Match 38.6%; Score 489.2; DB 6; Length 1378;
Best Local Similarity 65.0%; Pred. No. 2.5e-135;
Matches 722; Conservative 0; Mismatches 388; Indels 0; Gaps 0;

```

```

Db 496 ATGACCTTCCAGCCCAATATGACTTCAATGTAAGAAAGAACTGGAGAGAAAGCTGCACT 555
Qy 545 atgtcgctatctcaagggccaccacatggcttatgtcatcttccaccatgccaagc 604
Db 556 ATGTTGGTCATTCGCCAGGCGACACCAATGATGTTATTTATGCTTTCTACAAATCCAAAGC 615
Qy 605 tggctcagaagaatcaaatgtatcttcttgaagccatagccactgttaagcatgca 664
Db 616 TGGCTGAAGAAGATCAAAACCTTTATGCAATTAGCTCCAGCTGGCCACTGTAAGTATACA 675
Qy 665 aaagcccgaggccaatttctgtcgcagatagatgacatgaagagatgttggtgca 724
Db 676 AAAGCCTGTAAACAACCTAGTTATTCCTCCAAACCATCTGAAGTTATATATGCTG 735
Qy 725 aaaaagaattctgtatcagaccgaattctcagcaactgttatcttacccttggcc 784
Db 736 ACAAATATTTCTACCCACACAAATTTCTTGATCAATTTCTCCACCTCAAGTGTCTCC 795
Qy 785 agtgatctctgacagagttgtagtaataatcatatcatgttactctgggtgattcaacacca 844
Db 796 GTGAGACACTGATGATGATTTGTCACAGCATGCTTATTTATATCTTGTGGCTTGGACAGC 855
Qy 845 acaatatgacatgagccgagcgaagtgtatataatgctccacacactctgtggaacatctg 904
Db 856 CAACTTGAAACATGAGTGCGCTTGATGTGACGTGCAATATCCGAGAGAACTTACG 915
Qy 905 tgcataataatctacactgagagccagcagtgatgaatcttggtgaactccgggcatctgct 964
Db 916 TTCAAAACATGCTGCACCTGACCCAGCGTGTTAAATCTGGAATTTTCAAGCTTTTAAAT 975
Qy 965 gggggagtgagacccaataatcgtgaaatgacatcaacacactctgtaagtcagag 1024
Db 976 GGGGAAGTCCACCTGAGATGAGTGACCTTAATTCAGCCACACCTCCCTACTCAATG 1035
Qy 1025 tcaagatatgacggtccctacagcaatgtgagcagagtgctcagactggttcaaatc 1084
Db 1036 TGACCGCCATGATGATGCGCAATTTGCAGATGTGAGTGAGTGCGATGCGCTTGAC 1095
Qy 1085 cagaagacgtgaaatgctctctctgagtgagtgacaactcatcacaagaatctc 1144
Db 1096 CCCAAGATGTTGACCTTTGCTCCAAACCTTTCAATCTCAATTCACCAAGAGGATTC 1155
Qy 1145 ctgaatggctcagtgatgataatcattcattgaggttgattggttctaccgctatgacaatg 1204
Db 1156 TTCCATATATATCTGATGATTTATCTGGGCAATGAATGCTCTCAAGAATTTTCAATG 1215
Qy 1205 aaatccatcctgagtcagcagagagaga 1234
Db 1216 AATATATTTCTATGATGCGCAAAAGATAAAA 1245

RESULT 7
LOCUS A26690 1378 bp DNA PAT 05-APR-1995
DEFINITION Precursor of rabbit gastric lipase (comp.).
ACCESSION A26690
VERSION A26690.1 GI:905030
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct.
artificial sequence.
REFERENCE 1 (bases 1 to 1378)
AUTHORS Benicourt,C., Blanchard,C. and Junien,J.L.
TITLE Recombinant gastric lipase from rabbit and pharmaceutical
compositions
JOURNAL Patent: EP 0542629-A 10 19-MAY-1993;
INSTITUT DE RECHERCHE JOUVEINALE
LOCATION/Qualifiers
FEATURES
SOURCE
1..1378
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT 418 a 276 c 287 g 397 t
ORIGIN

```





Db 423 GAACTTACTACTACACAGATTCAGTGAATCTTGCGCTTACGCTTATGAAATG 482  
 Qy 479 ctgaagttgacctccgcgcagtgataaacttatttgcagaanaacggccaggaanaa 538  
 Db 483 CTAAATATGACCTTCAGCCCAATCGATTCATGTAAGAAACATGACGAGAACG 542  
 Qy 539 tctatttgcgcctatcacagggaaccacatgagcttattgacttccaccatgc 598  
 Db 543 TACACTATGTGGCCATTCACAGGACACACCATTTGTTTATTTGCTTTCCACACATC 602  
 Qy 599 cagagctggtcagagaataacaaatgatttgcctttagcaccatagccactgtaac 658  
 Db 603 CCAGCTGGCTAAAGATATCAAAACCTTCTATGCTAGCTCCGTTGCCACTGTAAGT 662  
 Qy 659 atgcaaaaagcccgaggacaacttctgtctgcgcagatataatgataaggattgt 718  
 Db 663 ATACAAAAGCCCTTATTAACAACTTAGATTGTTCCCAATCCCTCTTCAAGTTTATAT 722  
 Qy 719 ttggcaaaaagaattctgtatcacagaccagatttccagacaactgttattacctt 778  
 Db 723 TTGGTGCACAAATATTTCTACCCACACAACTTTTGATCAATTTCTGCTACTGAAGTGT 782  
 Qy 779 gtggcaggtgattcttgcatacagaatttgcataatcatatgtaacttgcgtgagatca 838  
 Db 783 GCTCCCGTAGATCTGATCTCTTTCGACGATGCTTATATATTTATTTGTTGATTTG 842  
 Qy 839 acacacaataatgatacagtagccgcagcaagtgtatgctgcgcacaactctgtcgaa 898  
 Db 843 ACACTAAGAACTTAAACACGAGTCTGTGATGTATCTATCAATATATACACAGAA 902  
 Qy 899 catctgcaaaaataattctacactgcagggcagcgaggaattccgtggaactccggcat 958  
 Db 903 CTTTGTTCACAAATATCTTCATTTGACCCAGGCTGTAACTCTGGGAAATTCACAGCTT 962  
 Qy 959 ttgacttggggagtagagaccaaacttgcgaataatgcaatcacgacaactcgttaagt 1018  
 Db 963 ATGACTGGGAGAGCCCACTTCAGATAGATGACATGATGACATGCCAACCTCCCTACT 1022  
 Qy 1019 acagagtcagagataagcaggtccctacagcaatgtgacagagagtcagagctgctt 1078  
 Db 1023 ACATATGACAGACCATTAATGACCAATGCGAGTGTGGAACGGTGGACAGACCTGTTGG 1082  
 Qy 1079 caaatccagaagaagtgtaaaatgctgctcctgcgtgagtgacacaactcatcataga 1138  
 Db 1083 CTGACCCCAAGATGTGGCTTTTGTTCCTCAAAACCTCCCAATCTATTATCCACAGG 1142  
 Qy 1139 aatctcgtgaatgggtcagcgaggttcatctgcgtggttgcagtcctccacagctatgt 1198  
 Db 1143 AGATTCTTTTACATCACTTGAGCTTTTATCTGGCAATGATGCCCCCTCAGAGAGTTT 1202  
 Qy 1199 acaatgaatacatcatcatgtat 1220  
 Db 1203 ACAATGACATTGTTTCTATGAT 1224  
 RESULT 10  
 A12714 1367 bp DNA PAT 09-AUG-1994  
 LOCUS A12714  
 DEFINITION Pregnastic lipase.  
 ACCESSION A12714  
 VERSION A12714.1 GI:579560  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1367)  
 AUTHORS PRECURSOR POLYPEPTIDE, DNA SEQUENCE CODING THEREFOR, VECTORS, HOST  
 TITLE ORGANISMS, AND PROCESSES INVOLVING SAME  
 JOURNAL Patent: WO 8603778-A 8 03-JUL-1986;  
 FEATURES Location/Qualifiers

source 1..1367  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 CDS 47..1243  
 /codon\_start=1  
 /product="pregnastic lipase"  
 /protein\_id="CAA01053.1"  
 /db\_xref="GI:579561"  
 /db\_xref="SWISS-PROT:P07098"  
 /translation="MMLLTMASLISVLGTHGLFGKLFHGSPEVTNISMITYWGY  
 PNEYEVVTEEDGYILEVNRIPYKKNKNGNNGORVPELQGLASATNMTSNIPNSL  
 AFLIADGYVWMLGNSRGNWARMNLVSDSEVAFSEDMKXDPATPIEYVK  
 TGOKOLHYVHSOGTTGFIATFSPNPLARKITFYALAVATYKYSILNLRFP  
 QSLPKFTFGKITYPHNFPQPLATEVCSKEMNLCSNMLFICGDSKNPTSRD  
 VYLSHPAGTSVQNMFWTOAVSGKQYIAKQVAVNSPVOVNRHVDOSOPTYVNRNVP  
 IAVNGSKDLADPODVGLLPLKPLNLIHKEIPFYNHLDFIWMADAPDEYNDIVSM  
 ISEDKK"  
 BASE COUNT 397 a 284 c 269 g 417 t  
 ORIGIN  
 Query Match 37.0%; Score 470; DB 6; Length 1367;  
 Best Local Similarity 64.2%; Pred. No. 1.4e-129;  
 Matches 707; Conservative 0; Mismatches 395; Indels 0; Gaps 0;  
 Db 119 ctgtgagaccagaagcatctatgataatagtaataatccacaatcaagctatccct 178  
 123 CTGGAACCCCTGAAGTGACTATGAACATTAATCTAGATGATTAATGTTGGGATACCAA 182  
 Qy 179 gtggagataatgagtgacacataagatggatatacccttctgttaacagagattctc 238  
 Db 183 ATGAGATATGAAGTTGTGCTGACGATGAGTTATTTCTTGAAGTCAATGAAATTCCTT 242  
 Qy 239 gaagcctagtgcaacttaagaagaacaggttccagcctgtgtgtaactgcagatggcc 298  
 Db 243 ATGGGAAGAAAATTCAGGGAATACAGGCCAGACACCGTTGTGTTTTCAGACATGTT 302  
 Qy 299 taattggaagtgtagcaactgtatccaaactgcgcacaataatagcctggtctatc 358  
 Db 303 TGTCTGCATCAGCCCAACAACTGAGTTTCCAACTGCGGACACAAAGCCCTTTCATTC 362  
 Qy 359 tggagagtgctggttttgaagtgtagtgagtggaacagagggagaaagccgtgctcgaa 418  
 Db 363 TGGCAGATGCTGTTATGATGTGTGGCTGGCAACAGCAGAGAAACCTGGGCCAGAA 422  
 Qy 419 aacacaagaacactccatagacacaagatgagtcctggtcctcagtatagatgagtg 478  
 Db 423 GAACTTGTACTATTACACAGATTCAGTGAATCTTGCGCTTACGCTTGTATGAAATG 482  
 Qy 479 ctgaagttgacctccgcgcagtgataaacttatttgcagaanaacggccaggaanaa 538  
 Db 483 CTAAATATGACCTTCAGCCCAATCGATTCATGTAAGAAACATGACGAGAACG 542  
 Qy 539 tctatttgcgcctatcacagggaaccacatgagcttattgacttccaccatgc 598  
 Db 543 TACACTATGTGGCCATTCACAGGACACACCATTTGTTTATTTGCTTTCCACACATC 602  
 Qy 599 cagagctggtcagagaataacaaatgatttgcctttagcaccatagccactgtaac 658  
 Db 603 CCAGCTGGCTAAAGATATCAAAACCTTCTATGCTAGCTCCGTTGCCACTGTAAGT 662  
 Qy 659 atgcaaaaagcccgaggacaacttctgtctgcgcagatataatgataaggattgt 718  
 Db 663 ATACAAAAGCCCTTATTAACAACTTAGATTGTTCCCAATCCCTCTTCAAGTTTATAT 722  
 Qy 719 ttggcaaaaagaattctgtatcacagaccagatttccagacaactgttattacctt 778  
 Db 723 TTGGTGCACAAATATTTCTACCCACACAACTTTTGATCAATTTCTGCTACTGAAGTGT 782  
 Qy 779 gtggcaggtgattcttgcatacagaatttgcataatcatatgtaacttgcgtgagatca 838  
 Db 783 GCTCCCGTAGATCTGATCTCTTTCGACGATGCTTATATATTTGTTGATTTG 842



RESULT 12

RNLIIP 1355 bp mRNA ROD 30-MAR-1995

LOCUS Rat mRNA for lingual lipase.

DEFINITION X02309.1 GI:56595

ACCESSION X02309.1 GI:56595

VERSION glycoprotein; lipase.

KEYWORDS Norway rat.

SOURCE Rattus norvegicus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 1355)

AUTHORS Docherty,A.J., Bodmer,M.W., Angai,S., Vargier,R., Riviere,C., Lowe,P.A., Lyons,A., Emhage,J.S. and Harris,T.J.

TITLE Molecular cloning and nucleotide sequence of rat lingual lipase cDNA

JOURNAL Nucleic Acids Res. 13 (6), 1891-1903 (1985)

MEDLINE 85215587

COMMENT Data kindly reviewed (12-FEB-1986) by A.J.P. Docherty.

FEATURES

source Location/Qualifiers

1..1355

/organism="Rattus norvegicus"

/db\_xref="taxon:10116"

49..102

/product="put. signal peptide (aa -18 to -1)"

49..1236

/note="lipase precursor"

/codon\_start=1

/protein\_id="CAA26179.1"

/db\_xref="GI:56596"

/db\_xref="SWISS-PROT:P04634"

/translation="MWLLITSVISTREGANGLGKLPENPEANMISQITWGP  
CQETEVTEDEYILGVRIPIGKNNSEIGRPVYVYLQHLIASATNMIANLPNNSLA  
FMLADAGYDVLGNSRGNTWRSKNVYSPDSVEWMSFEDMAYDLPATINFTVOKT  
GOEKIHVHGSGQTIGFIAFSTNPTLAKRKETVYALAPVATVYTOPLKIEFIP  
FLFKMGKMFLEPHYFEDDELGTEVCSREVLIDLSNTLIFCGPDKNNVSRPDI  
YLGHNPAFTSVQDFLHNAVLGSKFOAFNMGSSONLHYNOKTPPYDVSAMTPVY  
AVWNGNDIADPDVAMLEPKLSNLEFKHETLAVNHLDFIMANDAPQEVNENISM  
AED"

mat\_peptide 103..1233

/product="lipase (aa 1-377)"

misc\_feature 505..507

/note="aa 153 pot. lipid binding site"

misc\_feature 1355

/note="polya site"

BASE COUNT 400 a 291 c 276 g 387 t 1 others

ORIGIN

Query Match 36.8%; Score 467.6; DB 10; Length 1355;

Best Local Similarity 63.5%; Pred. No. 7.6e-128;

Matches 713; Conservative 0; Mismatches 409; Indels 0; Gaps 0;

Db 362 TAGCAGCGCTGGCTATGATGTGGGGAACGTCGAGCAATATCATGTCACAGA 421

419 aacacaagaacctcccatagacaagaatgagttcgtggccttcagttatgagatgg 478

422 AAAATGTACTACTACACAGACACTGATTTGATTTGGGCTTTGACGTTGTAATGG 481

479 ctaggcttgacctccctcagtgataaacttatttcagagaagaacggccagagaaga 538

482 CTAAATATGACCTTCCCGCACAAATTAACCTTATTGTACAGAAATCGACAGAGA 541

539 tctattatgctggctatcacagggcaccacacatggccttattgcatllccaccatgc 598

542 TACACTATGTTGGTCTCTCTCAGGGCACCATATGTTGTTTCCTCTTCTTACAAATC 601

599 cagactggtctcagaataatcaaatgatttttgccttgagaccatcagcactgttaagc 658

602 CTACACTGGCCAAAAAATCAAGACGTTTATGCAATTAAGCTCAGTTCTACCGGAAGT 661

659 atgcaaaaagcccgagcaaatlttgctgcgcagatatgatatgataagagatgt 718

662 ATACCAAAAGTCCCTTGAAAAAGATTATTCATTCATCTTCAAGCTTAATGT 721

719 ttggcaaaaagaattctgatatcagaccagaattctcagaacaattgattaccctt 778

722 TTGGCAAGAAATGTTCCCTCCGCCACACCTACTTGTATGATGATTTCTTGACCGAAGTGT 781

779 gtggcaggtgattcttgatcagatttgagtaataatgatttactctggttgatcca 838

782 GCTACGGGAGGTTCTTAATCTTCTCTCAGACMACCTTATTCATCTTCTGTGATTTGG 841

839 acaccaacaatatgacatgagccgagcaagtgatattgctggccaccactctgttgaa 898

842 ACAAAATAAATTAATATGAGTCTTTTGATGTATCTAGGCAATATCCAGCAGGGA 901

899 catctgtgcaaaatattctacacttgagagccagtcagttatctgtgtaactccggcat 958

902 CATCTGTTCAAGACTTCTCCACTGGGACGCTGTTAGATGTTGGAAATTTCAAGCT 961

959 ttgactggggagatgagccaaaataatctgggaataatgcaatcagccaactctctgaagt 1018

962 TCAACTGGGGAAGCCCATCCAGAACATGTTACATTAACCAAGAAAGCGCTCTCGAAT 1021

1019 acaagtcagagatatgacggtgccctacagcaaatgttgacaggaagtcagactgtt 1078

1022 ATGATGTCACGCCATGACTGTGCCAGTTGCAAGTGGCAAGCGTGCATACATCCTGG 1081

1079 caaatccagaagacgtgaanaatgctctctctgagtgagtgacaacctcatcataaga 1138

1082 CTGATCCCCAAGATGTCCGCATGCTGCTCCCAAACTCTCCAAACCTCCTGTTCCATAAGG 1141

1139 atattccgaatggctcagtgattcactcgtgggtttgagatgctctcaccgtatgt 1198

1142 AGATTCTTGCTACATATCCTGACCTGACTTCATCTGGGCAATGATGATGCCCTCAAGAGTGT 1201

1199 acaatgaaatcatcatctgatatgacagcagagggaggaaccaacc 1240

1202 ACAATGAGATGATTTCCATGATGAGCAGAAAGCTAAAGAAATC 1243

RESULT 13

MLYACLY 2358 bp mRNA ROD 24-FEB-1995

LOCUS M. musculus (C57 Black/6X CBA) LAL mRNA for lysosomal acid lipase.

DEFINITION 231689

ACCESSION 231689.1 GI:4456670

VERSION LAL; lipase; lysosomal acid lipase.

KEYWORDS house mouse.

SOURCE Mus musculus.

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2358)

AUTHORS Du,H., Witte,D.P. and Grabowski,G.A.







```

SEMHRVITADGYIILOVYRIPRPHGNKNNANHLGQRPVYFIQHLGLSATNWSINLPSNLG
FILADAGYDWMIGNSRGNTWQDEHLYYSPPDEPMFASFDMAFYDLPSTDFILRRT
GQKRLHYVHSGOCTTIFIAFSTSPITAEKIKVYALAPVATVYTSLEPKLIIPIH
FLFKITIGDKMFTPHTELEDFLEWCKSREPLDVLCKNALFAITGDKNNMSRLDY
YIAHNPAQTSVQNTLHWROAVKSGKFOAFDMGAPYQNMHMHQTPPLYNLTANVPI
AVWSADNDLLADPOVDVFLSKLSNLIYHKEIPVYNHLDPIWADAOEYNEIYSIM
AEDRK"
mat_peptide
130.1263
/EC-number="3.1.1.3"
/function="triacylglycerol acylhydrolase"
/evidence="experimental"
/product="pregastric esterase"
polya_signal
1378.1385
polya_site
1396
BASE COUNT 408 a 306 c 276 g 427 t
ORIGIN
Query Match 36.1%; Score 458; DB 4; Length 1417;
Best Local Similarity 62.1%; Pred. No. 5.8e-126;
Matches 741; Conservative 0; Mismatches 450; Indels 3; Gaps 1;
QY 44 tgaagatggtctcgtatctcgtgtgagcgtatgctcagaagaatgugaatcagtcac 103
DB 68 TGAATAATGTGGGCTACTGTGAACAGTGTGTTCATCCACATGCTGGAATGCAATTTT 127
QY 104 atatgc---caactaaagctgtagccagaagcatlcatgatatagtgaaatcacc 160
DB 128 GTTTCCTTGGAAAATGTCTAAGAACCTGAAGCCAGATATGAATGTAGTCAAGATATT 187
QY 161 aacatcaagctatccctggtgaggaatatgaagtcgaactgaagaatggtatccctt 220
DB 188 CCTACTGGGGCTACCAAGATGAGATGATGAATTAAGTTCGATGGTTATATCTTC 247
QY 221 ctgttaacaggaatccctcgaagccctagtcgaacctagaagaagcttccagagcctgag 280
DB 248 AGGTCTATCGGATTCCTCTATGGAAGAATAATGCTAATCATTTAGCTCAGAGACCTGTTG 307
QY 281 tgttactgcaagatgagcctagttggaagtgctgaactggaatttccaactgccaaca 340
DB 308 TGTTCCTGACAGATGCTTCTTGATCAGCCAAACGATGATTTCCAACCTGCCAAGA 367
QY 341 atagcctggtctcatcttcgcagatgctgttcttgcagtgtagtggggaacagagag 400
DB 368 ACAGCCTGGGCTCTCTGCGAGATGCTGTTATGACGTGCTGGGGAACACAGAG 427
QY 401 gaaacgctgtctcgaaaaacagaacaccttcacatagaccagaatgagctctggcctt 460
DB 428 GAAACACTGGGCCAGAGAACTTTATATCTTACACAGACTCCCCGGATTCGGGCTT 487
QY 461 tcaagtaatgataagatgagctgagttgaccttcctcgcagtgataaacttatttgcaga 520
DB 488 TCAGCTTGTGATGAATGGCGGAATATGACCTTCATCTACATGATTTCACTTAAGCA 547
QY 521 aaacggcccaagaaagaatctatctatgctgcgtatctcaagggaaccacacagggctta 580
DB 548 GAACGAGACAGAAAGCTACACTATGTGGCATTCCCAAGGACACACCATTTGGTTTA 607
QY 581 ttgcatttccacacgacagagcgtgcagaaatacaaaatgfatlttgctttagcac 640
DB 608 TCGCCTTTTTCACAGTCCACATTGGCTGAAAAAATCPAAGTCTTCTATGCAATTAGCCC 667
QY 641 ccaatagcacatgtaagatgcaaaaagcccggaagccaatlttltgctgycagagata 700
DB 668 CAGTTTCCACAGTAGAAGACAGAGAGCTGTTTAACAACCTTGCACTTATCTCTCACT 727
QY 701 tgaagtaacagagatgttcttgcaaaaagaattctgcatacagaccagatcttccagac 760
DB 728 TCCCTCTCAAGATTATATTGGTGACAAAATCTTCTACCCACACACTTTTTFGAACAAT 787
QY 761 aactgttatctaccttggtgcagagtgatcttgcacagatttgtagtaataatcagtc 820
DB 788 TTCTTGGTGTGAATGTGCTCCGTGAGACACTGGATGTCTTGTGAAGAAATGCTTGT 847

```

```

QY 821 tactctggtgagattcaaccacaacatatgaacatgagccgagcaagtgtatagctg 880
DB 848 TTGCCATTACTGGAAGTTGACATTAATAAACTTCACATGATGCTGCTAGATGTGATATAG 907
QY 881 cccacacctctgctggaacacatctgtgcaaaaatctacactggaagccagcgagtaatt 940
DB 908 CACATTAATCCACAGAGAACTTCTGTCAAAACACCCCTCACCTGGAGACAGGCTGTAACT 967
QY 941 ctggtgaactccgggcatcttgaactggtggagtgagaccaaatactggaataatgcaatc 1000
DB 968 CTGGGAATTCCAAGCTTTTGTACTGGGAGGCCCATATCAGAACTTAATGCAATATCATC 1027
QY 1001 agccaactccctgtaagtagcagagatgagatatgacggtccctcaacagcaatgtagacg 1060
DB 1028 AGCCCACTCTCCCACTTACATTAATACAGCCCAATGAATGCCAATTCAGATGAGAGTG 1087
QY 1061 gaggtcaggaactggttccaatccagaagacgtgaaaatgctgctcctgaggtgacca 1120
DB 1088 CTGACAAATGACCTGTGGCTGACCTCAGAGATGTGACTTCTGCTTCAAAACTCTCTA 1147
QY 1121 accctatctacataagaatatctcgaatggtgctcagctgagttcatctggggttgg 1180
DB 1148 ATCTCATTTACACAGAGAAATTCCAATTAACATCACTTGACTTTATCTGGGCAATGG 1207
QY 1181 atgtccctcaccgtatgtacaaatgaaatcaccatctgtagcagcagagagaga 1234
DB 1208 ATGCACCTCAAGAACTTTACATGAATTAATGTTCTTGTATGCGCCGAAGACAAA 1261

```

Search completed: January 30, 2002, 22:09:16  
Job time: 6263 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2002, 21:08:53 ; Search time 166.02 Seconds

(without alignments)  
6553.100 Million cell updates/sec

Title: US-09-578-063-46

Perfect score: 1269

Sequence: 1 atgttggaacctgtcaag.....gacggtgtgagccgcatgtt 1269

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues 1861242

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*

- 1: /SID58/gcgdata/geneseq/geneseqn/NA1980.DAT.\*
- 2: /SID58/gcgdata/geneseq/geneseqn/NA1981.DAT.\*
- 3: /SID58/gcgdata/geneseq/geneseqn/NA1982.DAT.\*
- 4: /SID58/gcgdata/geneseq/geneseqn/NA1983.DAT.\*
- 5: /SID58/gcgdata/geneseq/geneseqn/NA1984.DAT.\*
- 6: /SID58/gcgdata/geneseq/geneseqn/NA1985.DAT.\*
- 7: /SID58/gcgdata/geneseq/geneseqn/NA1986.DAT.\*
- 8: /SID58/gcgdata/geneseq/geneseqn/NA1987.DAT.\*
- 9: /SID58/gcgdata/geneseq/geneseqn/NA1988.DAT.\*
- 10: /SID58/gcgdata/geneseq/geneseqn/NA1989.DAT.\*
- 11: /SID58/gcgdata/geneseq/geneseqn/NA1990.DAT.\*
- 12: /SID58/gcgdata/geneseq/geneseqn/NA1991.DAT.\*
- 13: /SID58/gcgdata/geneseq/geneseqn/NA1992.DAT.\*
- 14: /SID58/gcgdata/geneseq/geneseqn/NA1993.DAT.\*
- 15: /SID58/gcgdata/geneseq/geneseqn/NA1994.DAT.\*
- 16: /SID58/gcgdata/geneseq/geneseqn/NA1995.DAT.\*
- 17: /SID58/gcgdata/geneseq/geneseqn/NA1996.DAT.\*
- 18: /SID58/gcgdata/geneseq/geneseqn/NA1997.DAT.\*
- 19: /SID58/gcgdata/geneseq/geneseqn/NA1998.DAT.\*
- 20: /SID58/gcgdata/geneseq/geneseqn/NA1999.DAT.\*
- 21: /SID58/gcgdata/geneseq/geneseqn/NA2000.DAT.\*
- 22: /SID58/gcgdata/geneseq/geneseqn/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1269	100.0	1269	22	AAAF5132	Human TANGO 294 ORF
2	1269	100.0	22	22	AAAF5131	Human TANGO 294 ORF
3	526.2	41.5	2453	22	AAH02936	Human Shear stress
4	489.2	38.6	1378	14	AAQ04310	ROL precursor. Circular
5	470	37.0	1367	7	AAAG0685	Sequence encoding
6	470	37.0	1367	7	AAAG0566	Sequence encoding
7	470	37.0	1367	17	AAAT5916	Human gastric lipase
8	468.4	36.9	1695	22	AAH57420	Human stomach cell
9	464.4	36.6	1336	6	AAAG5085	Rat prelingual lip
10	444.2	35.0	1528	15	AAQ06838	Canine gastric lip
11	444.2	35.0	1528	17	AAAT5915	Dog gastric lipase

12	444.2	35.0	1531	15	AAAF5131 Human TANGO 294 ORF
13	286	22.5	699	22	AAAF5131 Human TANGO 294 ORF
14	286	22.5	1308	22	AAAF5131 Human TANGO 294 ORF
15	257.8	20.3	617	21	AAAF5131 Human TANGO 294 ORF
16	193.2	15.2	572	21	AAAF5131 Human TANGO 294 ORF
17	100.2	7.9	402	22	AAAF5131 Human TANGO 294 ORF
18	100	7.9	1718	21	AAAF5131 Human TANGO 294 ORF
19	89.4	7.0	1483	21	AAAF5131 Human TANGO 294 ORF
20	84.8	6.7	859	21	AAAF5131 Human TANGO 294 ORF
21	72	5.7	1438	21	AAAF5131 Human TANGO 294 ORF
22	68	5.4	936	22	AAAF5131 Human TANGO 294 ORF
23	68	5.4	936	22	AAAF5131 Human TANGO 294 ORF
24	68	5.4	936	22	AAAF5131 Human TANGO 294 ORF
25	68	5.4	936	22	AAAF5131 Human TANGO 294 ORF
26	68	5.4	936	22	AAAF5131 Human TANGO 294 ORF
27	66.2	5.2	936	22	AAAF5131 Human TANGO 294 ORF
28	66.2	5.2	936	22	AAAF5131 Human TANGO 294 ORF
29	66.2	5.2	936	22	AAAF5131 Human TANGO 294 ORF
30	66.2	5.2	936	22	AAAF5131 Human TANGO 294 ORF
31	66.2	5.2	936	22	AAAF5131 Human TANGO 294 ORF
32	66.2	5.2	936	22	AAAF5131 Human TANGO 294 ORF
33	66.2	5.2	936	22	AAAF5131 Human TANGO 294 ORF
34	55.8	4.4	705	21	AAAF5131 Human TANGO 294 ORF
35	54	4.3	549	21	AAAF5131 Human TANGO 294 ORF
36	51	4.0	509	21	AAAF5131 Human TANGO 294 ORF
37	48.2	3.8	178	19	AAAF5131 Human TANGO 294 ORF
38	46.4	3.7	43676	21	AAAF5131 Human TANGO 294 ORF
39	38.6	3.0	1664976	19	AAAF5131 Human TANGO 294 ORF
40	36.8	2.9	13254	21	AAAF5131 Human TANGO 294 ORF
41	36.6	2.9	244	22	AAAF5131 Human TANGO 294 ORF
42	36.6	2.9	3271	19	AAAF5131 Human TANGO 294 ORF
43	36.6	2.9	3282	17	AAAF5131 Human TANGO 294 ORF
44	36.6	2.9	3282	21	AAAF5131 Human TANGO 294 ORF
45	36.6	2.9	3282	21	AAAF5131 Human TANGO 294 ORF

## ALIGNMENTS

AAAF5132	1	AAAF5132 standard; cDNA, 1269 BP.
AC	AAAF5132;	
XX	30-MAR-2001 (first entry)	
DE	Human TANGO 294 ORF.	
XX		
KW	Gene therapy; TANGO protein; INTERCEPT protein; neurological disorder;	Canine gastric lip
KW	central nervous system; focal brain disorder; bipolar affective disorder;	Human protein HP03
KW	global diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;	Human protein HP03
KW	senile dementia; Huntington's disease; amyotrophic lateral sclerosis;	Human secreted exp
KW	Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;	Human stomach spec
KW	neuropsychiatric; psychoactive substance use; anxiety; ss.	Soybean acid triac
XX		
OS	Homo sapiens.	Rice acid triacylg
XX		
PN	WO200077239-A2.	Corn acid triacylg
XX		
PD	21-DEC-2000.	Soybean acid triac
XX		
XX	4-MAY-2000; 2000WO-US14858.	Oligonucleotide D1
XX	14-JUN-1999; 99US-033159.	Oligonucleotide D1
XX		Oligonucleotide D1
XX	(MILL-) MILLENNIUM PHARM INC.	Oligonucleotide D1
XX		Oligonucleotide D1
XX	McCarthy SA, Fraser CC, Sharp JD, Barnes TM;	Oligonucleotide D1
XX	WPI: 2001-032313/04.	Oligonucleotide D1
DR	P-PSDB; AAB66065.	Oligonucleotide D1
XX		Oligonucleotide D1

PT TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for  
PT screening assays and diagnostic assays and for the treatment of  
PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's  
PT disease.

Claim 1; Fig 6; 359pp; English.

CC The present invention relates to TANGO or INTERCEPT proteins and coding  
CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB6031-B66037/  
CC AAB6064-B66083 and AAB6085). The TANGO/INTERCEPT proteins and coding  
CC sequences are useful for the treatment of neurological disorders such as  
CC central nervous system (CNS) disorders, CNS-related disorders, focal  
CC brain disorders, global-diffuse cerebral disorders and other  
CC neurological and cerebrovascular disorders. The CNS disorders include  
CC Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic  
CC lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,  
CC autonomic function disorders such as hypertension and sleep disorders,  
CC neuropsychiatric disorders, psychoactive substance use disorders,  
CC anxiety, and bipolar affective disorder.

XX Sequence 1269 BP; 358 A; 268 C; 309 G; 334 T; 0 other:

Query Match 100.0%; Score 1269; DB 22; Length 1269;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgttggaacctgttcagaacagtgatgtgttcacacagaatgaaatgtgcttcg 60  
Db 1 atgttggaacctgttcagaacagtgatgtgttcacacagaatgaaatgtgcttcg 60  
QY 61 attctggggtggtatatttccagaagaatgtgaattcagtaacatgtccaactaaagct 120  
Db 61 attctggggtggtatatttccagaagaatgtgaattcagtaacatgtccaactaaagct 120  
QY 121 gtggaccagagacatcatgaatattagttgaatcatccacatcaagggtatcccggt 180  
Db 121 gtggaccagagacatcatgaatattagttgaatcatccacatcaagggtatcccggt 180  
QY 181 gaggaataatgaagtcgcgaactgaagatgtgtatataccttctgttaacagatccctga 240  
Db 181 gaggaataatgaagtcgcgaactgaagatgtgtatataccttctgttaacagatccctga 240  
QY 241 ggcttagtgcgaactaagaagaaggttcgaagcctgtgtgttactgtcagatgacctga 300  
Db 241 ggcttagtgcgaactaagaagaaggttcgaagcctgtgtgttactgtcagatgacctga 300  
QY 301 gttagaggtgttagcaactgattccaactgcccacaataagcctggggttattctgt 360  
Db 301 gttagaggtgttagcaactgattccaactgcccacaataagcctggggttattctgt 360  
QY 361 gcaagatcgtgttttgacgtgtgagtgagacagcagggaagcctgtgttcgaaaa 420  
Db 361 gcaagatcgtgttttgacgtgtgagtgagacagcagggaagcctgtgttcgaaaa 420  
QY 421 cacaagaacacttccatagacaaagatgagttcctgggcttcaatgtatgagatgagct 480  
Db 421 cacaagaacacttccatagacaaagatgagttcctgggcttcaatgtatgagatgagct 480  
QY 481 aggtttgaacctctcgtcagtgatataaacttttttgcagaaaaagggccaggaataaact 540  
Db 481 aggtttgaacctctcgtcagtgatataaacttttttgcagaaaaagggccaggaataaact 540  
QY 541 tattatgtcggctatccacagggcaccacatgggcttattgtcattttccaccatgcca 600  
Db 541 tattatgtcggctatccacagggcaccacatgggcttattgtcattttccaccatgcca 600  
QY 601 gagctggctcagaataatcaaatgtattgtctttagcaccatagccactgttaagcat 660  
Db 601 gagctggctcagaataatcaaatgtattgtctttagcaccatagccactgttaagcat 660  
QY 661 gcaaaaagcccgagacaaattttgtctgcagatatgatatgacaggagattttt 720  
Db 661 gcaaaaagcccgagacaaattttgtctgcagatatgatatgacaggagattttt 720

Db 661 gcaaaaagcccgagacaaattttgtctgcagatatgatatgacaggagattttt 720  
QY 721 ggcacaaaagaattctgtatcagacagatcttcagacaaactgttattacattgtc 780  
Db 721 ggcacaaaagaattctgtatcagacagatcttcagacaaactgttattacattgtc 780  
QY 781 ggcaggtgattcttgcataatttgaatattatcatgttaactctcgggtgattaac 840  
Db 781 ggcaggtgattcttgcataatttgaatattatcatgttaactctcgggtgattaac 840  
QY 841 accacacatatgaacatgagccggcagagtgatatactcgcacacatctgtcggagaca 900  
Db 841 accacacatatgaacatgagccggcagagtgatatactcgcacacatctgtcggagaca 900  
QY 901 tctgtgcaaaatattctaacactgtgagccagcgagtgaaattctgtgtaactccggcatt 960  
Db 901 tctgtgcaaaatattctaacactgtgagccagcgagtgaaattctgtgtaactccggcatt 960  
QY 961 gactgggggtgtgagaccacaaaatcttgaaaaaatgacaatacagccaactcctgtgaagtac 1020  
Db 961 gactgggggtgtgagaccacaaaatcttgaaaaaatgacaatacagccaactcctgtgaagtac 1020  
QY 1021 agagtcagagatatgacggttcctcaccagcaatgtggaacagaggttcaggactgcttca 1080  
Db 1021 agagtcagagatatgacggttcctcaccagcaatgtggaacagaggttcaggactgcttca 1080  
QY 1081 aatccagaagaagcgtgaaaatgtctctctgtgaggttgacaaactcatcacaataaagat 1140  
Db 1081 aatccagaagaagcgtgaaaatgtctctctgtgaggttgacaaactcatcacaataaagat 1140  
QY 1141 attcctaaatgggtcaccggtgattcatctcgtgggttgatgactcctcaccgatactac 1200  
Db 1141 attcctaaatgggtcaccggtgattcatctcgtgggttgatgactcctcaccgatactac 1200  
QY 1201 aatgaataatcatcctcgtatgacagcagagagagacacacttccagggagcgtgtgag 1260  
Db 1201 aatgaataatcatcctcgtatgacagcagagagagagacacacttccagggagcgtgtgag 1260  
QY 1261 gccgtattg 1269  
Db 1261 gccgtattg 1269

RESULT 2  
AAF45131  
ID AAF45131 standard; cDNA; 2044 BP.  
XX  
AC AAF45131;  
XX  
DT 30-MAR-2001 (first entry)  
XX  
DE Human TANGO 294 cDNA.  
XX  
KW Gene therapy; TANGO protein; INTERCEPT protein; neurological disorder;  
KW central nervous system; focal brain disorder; bipolar affective disorder;  
KW global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;  
KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;  
KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;  
KW neuropsychiatric; psychoactive substance use; anxiety; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200077239-A2.  
XX  
PD 21-DEC-2000.  
XX  
PF 24-MAY-2000; 2000WO-US14858.  
XX  
PR 14-JUN-1999; 99US-0333159.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI McCarthy SA, Fraser CC, Sharp JD, Barnes TM;

XX WPI: 2001-032313/04.  
 DR P-PSDB: AAB66065.  
 XX TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for  
 PT screening assays and diagnostic assays and for the treatment of  
 PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's  
 PT disease.  
 XX  
 XX Claim 1; Fig 6; 359pp; English.  
 PS  
 XX The present invention relates to TANGO or INTERCEPT proteins and coding  
 CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,  
 CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding  
 CC sequences are useful for the treatment of neurological disorders such as  
 CC central nervous system (CNS) disorders, CNS-related disorders, focal  
 CC brain disorders, global-diffuse cerebral disorders and other  
 CC neurological and cerebrovascular disorders. The CNS disorders include  
 CC Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic  
 CC lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,  
 CC autonomic function disorders such as hypertension and sleep disorders,  
 CC neuropsychiatric disorders, psychoactive substance use disorders,  
 CC anxiety, and bipolar affective disorder.  
 XX  
 SQ Sequence 2044 BP; 618 A; 401 C; 460 G; 565 T; 0 other:  
 Query Match 100.0%; Score 1269; DB 22; Length 2044;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgttggaacctgtgcaagacgtggtgtcttcacacagaatggaatgtgcttctg 60  
 Db 126 atgttggaacctgtgcaagacgtggtgtcttcacacagaatggaatgtgcttctg 185  
 QY 61 attctggggcgtatatttcacagaaatgtgaattcagatcatatgccactaaagt 120  
 Db 186 attctggggcgtatatttcacagaaatgtgaattcagatcatatgccactaaagt 245  
 QY 121 gtggaaccagaacatcatgaataattgtaaatcatcacaacatcaaggctatccgt 180  
 Db 246 gtggaaccagaacatcatgaataattgtaaatcatcacaacatcaaggctatccgt 305  
 QY 181 gaggaattgaaatgcgaacatgaatgtggtatattccttctgttaacagattctctga 240  
 Db 306 gaggaattgaaatgcgaacatgaatgtggtatattccttctgttaacagattctctga 365  
 QY 241 ggctagtgcaacctaaagaacagaggtccaggcctgtgtgttactgacatgtgctta 300  
 Db 366 ggctagtgcaacctaaagaacagaggtccaggcctgtgtgttactgacatgtgctta 425  
 QY 301 gtggaagtgctgaactgtgattccaaactgtcccaacaatagcctggcttcatctg 360  
 Db 426 gtggaagtgctgaactgtgattccaaactgtcccaacaatagcctggcttcatctg 485  
 QY 361 gcaaatgtgttttggagtgttgatgggaacaggggaaagcctgtgtctgaaaa 420  
 Db 486 gcaaatgtgttttggagtgttgatgggaacaggggaaagcctgtgtctgaaaa 545  
 QY 421 cacaagaacctccatagacacaaagatagttctgggcttcagttatgatagtatgct 480  
 Db 546 cacaagaacctccatagacacaaagatagttctgggcttcagttatgatagtatgct 605  
 QY 481 aggttgacctctcctgcagtgataaacttatttgcagaaaacggccaggaagaatc 540  
 Db 606 aggttgacctctcctgcagtgataaacttatttgcagaaaacggccaggaagaatc 665  
 QY 541 tatatgctggcctttcaccagggacacacacacacacacacacacacacacacacac 600  
 Db 666 tatatgctggcctttcaccagggacacacacacacacacacacacacacacacacac 725  
 QY 601 gactgtgctcagaataatgatatttctttagaccacatagcactgttaagat 660  
 Db 726 gactgtgctcagaataatgatatttctttagaccacatagcactgttaagat 785

Db 726 gactgtgctcagaataatgatatttctttagaccacatagcactgttaagat 785  
 QY 661 gcaaaaagcccgagaccacaaatttctgtctgccagatcatgatcaaggatgtctt 720  
 Db 786 gcaaaaagcccgagaccacaaatttctgtctgccagatcatgatcaaggatgtctt 845  
 QY 721 ggcacaaaagatttctgtatcagaccagattcttcacagaaactgttatttacttctgt 780  
 Db 846 ggcacaaaagatttctgtatcagaccagattcttcacagaaactgttatttacttctgt 905  
 QY 781 ggcacagttattctgtatcagaccagatttctgtatcagaccagatttctgtatcagacc 840  
 Db 906 ggcacagttattctgtatcagaccagatttctgtatcagaccagatttctgtatcagacc 965  
 QY 841 accaacaatgatgaatagcccgagcaagtgtatatgtctgccacactctgtctggaaca 900  
 Db 966 accaacaatgatgaatagcccgagcaagtgtatatgtctgccacactctgtctggaaca 1025  
 QY 901 tctgtgcaaaatattcttactgagccagcagtgaaattctgtgtaactccgggcatctt 960  
 Db 1026 tctgtgcaaaatattcttactgagccagcagtgaaattctgtgtaactccgggcatctt 1085  
 QY 961 gactggggagtgagaccacaaatctggaacaaatgcaatcgaactcctgttaagtac 1020  
 Db 1086 gactggggagtgagaccacaaatctggaacaaatgcaatcgaactcctgttaagtac 1145  
 QY 1021 agagtcagagatgatagcggctcccaagaacatgtggacaggaagtcaggactgtctta 1080  
 Db 1146 agagtcagagatgatagcggctcccaagaacatgtggacaggaagtcaggactgtctta 1205  
 QY 1081 aatccagaagacggtgaaatattctgtctgtgagtgaccacccatctactacaagaat 1140  
 Db 1206 aatccagaagacggtgaaatattctgtctgtgagtgaccacccatctactacaagaat 1265  
 QY 1141 attcctgattggctcagctgatttcatctcgtgggtttgagtgctccaccagatgtac 1200  
 Db 1266 attcctgattggctcagctgatttcatctcgtgggtttgagtgctccaccagatgtac 1325  
 QY 1201 aatgaatcatcatctgtatgtcagcaggaaggagacacacatttccaggaaggtgtgag 1260  
 Db 1326 aatgaatcatcatctgtatgtcagcaggaaggagacacacatttccaggaaggtgtgag 1385  
 QY 1261 gccgtattg 1269  
 Db 1386 gccgtattg 1394

RESULT 3  
 AAH02906  
 ID AAH02906 standard; DNA; 2493 BP.  
 XX AAH02906;  
 AC  
 XX  
 XX 15-JUN-2001 (first entry)  
 DT  
 XX  
 DE Human shear stress-response coding sequence SEQ ID NO: 65.  
 XX  
 KW Human; shear stress-response protein; vascular disease;  
 OS arteriosclerosis; ds.  
 XX  
 XX Homo sapiens.  
 XX  
 XX MO200125427-A1.  
 XX  
 XX 12-APR-2001.  
 PD  
 XX  
 PF 02-OCT-2000; 2000MO-JP06840.  
 XX  
 XX 01-OCT-1999; 99JP-0280976.  
 PR  
 XX  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 PA  
 PA (NOJIMA ) NOJIMA H.  
 XX



RESULT	5	
AA60685		
ID	AA60685	standard; DNA; 1367 BP.
XX		
AC	AA60685;	
XX		
DT	08-JUN-1991	(first entry)
XX		
DE	Sequence encoding pregastric lipase.	
XX		
KW	Precursor polypeptide; secretion vector; enzyme; ss.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	sig_peptide	47..103
FT		/*tag= a
FT	mat_peptide	104..1243
FT		/*tag= b
XX		
PN	W08603778-A.	
XX		
PD	03-JUL-1986.	
XX		
PE	23-DEC-1985;	85WO-GB00599.
XX		
PR	21-DEC-1984;	84GB-0032483.
PR	23-DEC-1985;	85MO-GB00599.
PR	01-JAN-1986;	86GB-0019568.
XX		
PA	(BREW-) BREWING RES FOUND.	
PA	(TUBB/) TUBB R S.	
XX		
P1	Tubb RS;	
XX		
DR	WPI: 1986-162910/28.	
DR	P-PSDB; AAP60724.	
XX		
XX	New precursor polypeptide of defined sequence - and corresp. DNA	
PT	used to transform hosts for prodn. of the polypeptide	
PS	Example; Fig 5; 60pp; English.	
XX		

The inventors claim a DNA sequence encoding AAP60723 linked to a promoter upstream and a gene for a polypeptide downstream. Particular examples are the yeast enzyme AWG, the mammalian enzyme gastric lipase and the mammalian lymphokine, interferon- $\alpha$ 1n2.

XX Sequence 1367 BP; 397 A; 284 C; 269 G; 417 T; 0 other:  
SQ

Query Match 37.0%; Score 470; DB 7; Length 1367;  
Best Local Similarity 64.2%; Pred. No. 4,8e-139;  
Matches 707; Conservative 0; Mismatches 395; Indels 0; Gaps 0;

QY 119 ctgtgaccacgaagcattcatgataatgtaaatcaatcaacatcaaggctacc 178  
DB 123 ctggaagccctgaagtgactatgacattgacatgataatgattcttggatacc 182  
QY 179 gtaggaataatgagtcgcaacgcaagatgagatgataatcccttcgttaacagatcc 238  
DB 183 atgaagaataatgagtgagtcgcaagatgagatgataatcccttcgttaacagatcc 242  
QY 239 gaggcctagtgcaacctaagaagacaggtccagcctgtgtgttactgacatgccc 298  
DB 243 atggagaagaanaattcagggaaatcacagcagagacctgtgtgttcttgcagatggt 302  
QY 299 tagttgaggtgtgtagcaactgattccacactgcccacaataagcctgggtctatc 358  
DB 303 tgccttcacagccacaanaacttgatccaaactgcccgaacacagcctgtgcttctc 362  
QY 359 tggcagatgctggttttgacgtgtgtgagtgaggacagcaggggaacgcctgttcgaa 418  
DB 363 tggcagatgctggttttgatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 422  
QY 419 aacacaagaacacttcacatagaccagaatgagttcctgtggtctcattatgatatg 478  
DB 423 gaactctgtactatctaccagatctgaaattcctgtgtgtgtgtgtgtgtgtgtgtgt 482  
QY 479 ctgagttgacctctgtagatgataaacttttttgcgaaagaagggccagaagaaga 538  
DB 483 ctaaaatgacctccagacacaactgactctctgtaagaanaactgacagaagcagc 542  
QY 539 tctatattctgcgtatctacacagcaccacactgagccttattgtcattctccacatgc 598  
DB 543 taacatattgttgacatctccacagcaccacactgtgtttattgtcttctccacatc 602  
QY 599 cagaagctgtgctcagaataatcaaatgtatttgccttgaacccaatagccaactgtta 658  
DB 603 ccagcctgtgctcagaataatcaaaccttctatgtctcgtcctgtgtcgaactgtga 662  
QY 659 atgcaaaaagcccgagcaaaaattttgtgtgtgcagataagatgaagatgtgt 718  
DB 663 atcaaaaagccctataaacaacttaagattgttcccaatccctcccaagttcata 722  
QY 719 ttgcaaaaagaatttctgtatcagaccagattctcagacaactgttattacatt 778  
DB 723 ttggtgacaaaatattctacccacacacttcttgatcaattcttgcctgaagatgt 782  
QY 779 gtggcaggtgtatctctgacagattgtgataatcatatgttaactctgtgtgtgtca 838  
DB 783 gctccgtgtagatgtcgaattcttcttgcagcatctcttattataatttggatgtg 842  
QY 839 acacacaatattgacaagcagcagcaggtatatactgtgcacacactctgtcggaga 898  
DB 843 acagtaagaacttaacacagagtcgtgtgtgtatctatcacaataatccacagga 902  
QY 899 catctgtgcaaaaatcttctacactgtgagcagcaggtgaattctgtgtactccggac 958  
DB 903 ctctgttcaaaaatcattgtccalttgaccagcgtgttaagtctgtgaaatccagact 962  
QY 959 ttgactgtgggtgtgtgagccaaaatctgaaaaatgtcaatcagccaactcctgtta 1018  
DB 963 atgactgtgggtgtgtgagccaaaatctgaaaaatgtcaatcagccaactcctgtta 1022  
QY 1019 acagagctgagagatagagcctccacagcagatgtgagcagaggttcagagctgtctt 1078  
DB 1023 acaatgtgcagccatgaatgttaccatattgcagtggtggaagcgtgtgtgtg 1082  
QY 1079 caaatccagaagacgtgaaatgtctctctgtgaggtgacaaacccatctaccataga 1138

DB 1083 ctgaccocccagaatgtgtgctcttctgtcccaaaactccccaatctatttaccacaag 1142  
QY 1139 atattcctgaatgtgctcagctgagatctcatctgggtgtgtgagctcctcaccgta 1198  
DB 1143 agattccttcttaccacacttgacttatactcgtggcaatgagtcgcccagaagatt 1202  
QY 1199 acatgaataatccatctcatg 1220  
DB 1203 acaatgacattgttctatgat 1224

## RESULT 6

AA060566  
ID AA060566 standard; DNA; 1367 BP.

AA060566;  
XX

22-AUG-1991 (first entry)  
XX

DE Sequence encoding human pregastric lipase.  
XX

KW Cystic fibrosis therapy; enzyme; lipase deficiency; ss.  
XX

OS Homo sapiens.  
XX

FT Key Location/Qualifiers  
FT CDS 47..103  
FT mat\_peptide 104..1243  
FT /\*tag= b  
XX

PN W08601532-A.  
XX

PD 13-MAR-1986.  
XX

PF 15-AUG-1985; 85WO-GB00364.  
XX

PR 21-AUG-1984; 84GB-0021210.  
XX

PR 15-AUG-1985; 85WO-GB00364.  
XX

PR 01-JAN-1986; 86GB-0008897.  
XX

PA (CELL-) CELTECH LTD.  
XX

PA (LOWE/) LOWE P A.  
XX

PI Lowe PA.  
XX

DR WPI; 1986-081634/12.  
XX

DR P-PSDB; AAP60658.  
XX

PT New gastric lipase protein, esp. of human origin - for treating  
XX lipase deficiency, and DNA sequences coding for it

PS Disclosure; Fig 3; 39p; English.  
XX

CC The inventors claim a pregastric lipase protein and a gene encoding  
XX it. Gastric lipase protein is useful for oral administration to

CC treat lipase deficiency, e.g. cystic fibrosis or pancreatitis.  
XX

SQ Sequence 1367 BP; 397 A; 284 C; 269 G; 417 T; 0 other:

Query Match 37.0%; Score 470; DB 7; Length 1367;  
Best Local Similarity 64.2%; Pred. No. 4,8e-139;  
Matches 707; Conservative 0; Mismatches 395; Indels 0; Gaps 0;

QY 119 ctgtgaccacgaagcattcatgataatgtaaatcaatcaacatcaaggctacc 178  
DB 123 ctggaagccctgaagtgactatgacattgacatgataatgattcttggatacc 182  
QY 179 gtaggaataatgagtcgcaacgcaagatgagatgataatcccttcgttaacagatcc 238  
DB 183 atgaagaataatgagtgagtcgcaagatgagatgataatcccttcgttaacagatcc 242



QY 239 gagcgctagtgaacctaagaacaggttcacagcgctgtgtgttactgcagcatgccc 298  
 DB 243 atggaagaagaaatcaggaataacagagccagagcctgtgttcttcagcagatggtt 302  
 QY 299 taattggaggtgctagcagcaggtattccaaacctgcacacatgagcctggtctcattc 358  
 DB 303 tgccttgatcgcagcacaacactgatttccaaacctgcagacacagcgtctgtctcattc 362  
 QY 359 tggcagatgctggttttgcagctgtgtatggggaacagcaggggaaacgctgtgtcgaa 418  
 DB 363 tggcagatgctggtttatgtgtgtgtgtgcagacagcaggaagaaacacctgtgcagaa 422  
 QY 419 aacacaaacacacctcccttagacccaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 478  
 DB 423 gaacattgtactatcacaccagattcagatttctgtgtgtgtgtgtgtgtgtgtgtgt 482  
 QY 479 ctagggttgacctcttcctcagtgatgaacttatttgcagaaacagggccaggaagaa 538  
 DB 483 ctaaatatgacctctccagccacacatcgactcattgttaagaagaaacctgtgacgaagc 542  
 QY 539 tctatattgcgtctattcacagggcacccatgtggtcttatttgcatttccacacatgc 598  
 DB 543 tactactatgttgccattccagggcacccatgtggtcttatttgcatttccacacatgc 602  
 QY 599 cagagctgtgctcagaacaaatcgtatttctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 658  
 DB 603 cagagctgtgctcagaacaaatcgtatttctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 662  
 QY 659 atgcacaaagcccgagcccaatttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 718  
 DB 663 atacaaagcccttaacaaacttgaattgttctcctcaacccctcctcaagttatatt 722  
 QY 719 ttggcaaaaagaattcttctgtatcagaccagattctcagacatttctgtatttaccctt 778  
 DB 723 ttgtgtgacaaataattcttaccacacacattcttgcataatttctgtactatgagtg 782  
 QY 779 gtggcaggtgattcttgcacagattgtgtatataatcagttacttctgtgtgtgtatca 838  
 DB 783 gctcccgagatgtgtatgtcttcttgcagacatgcttatttatttatttgtgtattgt 842  
 QY 839 acacacaaatgtacatgagccgagcaagtgtatgtgtgtgtgtgtgtgtgtgtgtgtgt 898  
 DB 843 acagtaagaacttaccacagagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 902  
 QY 899 catcgtgtcaaaatattcttaccctgtgagccaggtgagttctgtgtgtgtgtgtgtgt 958  
 DB 903 ctctgtgtcaaaatattcttaccctgtgagccaggtgagttctgtgtgtgtgtgtgtgt 962  
 QY 959 ttgactggggaggtgagcacaacaaatcctggaacaaatgcacacagcgaactcctgtgaagt 1018  
 DB 963 atgactggggaggtgagcagcttgcagatagatgactatgactatgactatgactatgact 1022  
 QY 1019 acagatcagagatgacagctgtcctacagcaatgtgtgacaggaaggtcagagctgtt 1078  
 DB 1023 acaatgttacgcagctgatttaccacatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1082  
 QY 1079 caaatcaggaagctgtgaaatgctgtcctcaggtgtgaggtgacacacacacacacaa 1138  
 DB 1083 ctgacccccaagatgt 1142  
 QY 1139 atattcctgaatgggtcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1198  
 DB 1143 agattccttttacaatcactgtgacttattcctgtgtgtgtgtgtgtgtgtgtgtgtgt 1202  
 QY 1199 acaatgaataatcatcatctgtat 1220  
 DB 1203 acaatgacattgttctatgtat 1224

RESULT 7  
 AAT58916  
 ID AAT58916 standard; DNA; 1367 BP.

XX AC AAT58916;  
 XX DT 19-AUG-1997 (first entry)  
 XX DE Human gastric lipase coding sequence.  
 XX KW Duodenal; gastric; lipase; transgenic; plant; recombinant; extract; ss;  
 XX KW food; absorption; fat; pancreatic; cystic fibrosis; exocrine; dairy;  
 XX KW hydrolysis; trans-esterification; substrate; enzyme; biofuel; human.  
 XX OS Homo sapiens.  
 XX FH Key location/Qualifiers  
 XX FT CDS 47..1243  
 XX FT /\*tag= a  
 XX FT /product= gastric lipase  
 XX FT sig\_peptide 47..103  
 XX FT /\*tag= b  
 XX FT mat\_peptide 104..1240  
 XX FT /\*tag= c  
 XX PN W09633277-A2.  
 XX PD 24-OCT-1996.  
 XX PE 19-APR-1996; 96WO-FR00606.  
 XX PR 20-APR-1995; 95FR-0004754.  
 XX PA (BIOC-) BIOCEM SA.  
 XX PA (LJOU ) INST RECH JOUVEINAL.  
 XX PI Baudino S, Benicourt C, Cudrey C, Gruber V, Lenee P;  
 XX PI Merot B;  
 XX DR WPI: 1996-485783/48.  
 XX DR P-PSDB: AAM09383.  
 XX PT Recombinant human or canine pre-duodenal lipase proth. in transgenic  
 XX PT plants - useful for facilitating absorption of fat, as bio-catalysts  
 XX PT and for prodn. of bio-fuel  
 XX PS Claim 3; Fig 4; 130pp; French.  
 XX XX  
 CC This is the nucleotide sequence encoding the human pre-duodenal (1.e.  
 CC gastric) lipase enzyme. The sequence can be used to generate transgenic  
 CC plants producing recombinant lipase in an enzymatically active form.  
 CC CC Alternatively bases encoding amino acids 20-23 or 20-73 (1-4 or 1-54 of  
 CC the mature protein, respectively) can be deleted to form the derivatives  
 CC designated delta-4 or delta-54 respectively. Plants, or their extracts,  
 CC expressing the lipases or the truncated derivatives, can be used:  
 CC (a) as pharmaceuticals or food to facilitate absorption of fat, either  
 CC in healthy subjects or in patients with inadequate levels of  
 CC gastric/pancreatic lipase e.g. those with cystic fibrosis or exocrine  
 CC pancreatic insufficiency. In the elderly or in patients undergoing medical  
 CC treatment which alters fat adsorption; (b) for performing industrial or  
 CC agricultural reactions, e.g. in processing of fats or in the dairy  
 CC industry, for hydrolysis or transesterification reactions, etc., where  
 CC the plant material may provide both enzyme and substrate. The transgenic  
 CC plants can also be used for biofuel production.  
 CC XX  
 SQ Sequence 1367 BP; 397 A; 284 C; 269 G; 417 T; 0 other;

Query Match 37.0%; SCORE 470; DB 17; Length 1367;  
 Best Local Similarity 64.2%; Pred. No. 4.8e-139;  
 Matches 707; Conservative 0; Mismatches 395; Indels 0; Gaps 0;

QY 119 ctgtgaccagaagcatcatgaatattagtgaataatcacaatcaagctatccct 178  
 DB 123 ctggaagccctgaagtgactatgaacattatgtcagatgactattcgtgtgtgtgtgtgt 182

```

QY 179 gtgaggaataatgaagtcgaactgaagatgggtatattcttctgtttaacaggaattctc 238
DB 183 atgaagaataatgaagtcgtgactgaagatggttatattcttgaatcaatagaattctc 242
QY 239 gaggcctagtcgaagaagacaggttcacagcctgtgtgttaactgacagatgccc 298
DB 243 atgggaagaataatcagaggaatacagcagagacccgtgtgtttttgacagctggt 302
QY 299 taagtggagtggtctagcagactggaattccaaactgcccacaataagcctggctcattc 358
DB 303 tgcctgtcatcagccacaactggaattccaaactgcccacaataagcctgtctcattc 362
QY 359 tggcagagtcgtgttctgactgtgagtgaggggaacaggggaacccgtgctgaa 418
DB 363 tggcagagtcgtgttctgactgtgagtgaggggaacaggggaacccgtgctgaa 422
QY 419 aacaaagacactcccaataagacagaatgagtcgagcttccagtaatgatgagatg 478
DB 423 gaaactgtactattccacagattcagttgattcttggtcttccagcttgatgaatg 482
QY 479 ctgagttgaccttccctcagatgataaacttattctgcagaaaacgggccaagaaa 538
DB 483 cttaataatgacttccacacacacacacacacacacacacacacacacacacacac 542
QY 539 tctattatgctcgacttccacagggcaccacacacacacacacacacacacacacac 598
DB 543 taactatgctcgacttccacagggcaccacacacacacacacacacacacacacac 602
QY 599 cagaagctgctcagaaaatacaaatgtaatttgcctttagcaccacacacacacacac 658
DB 603 ccagcctgctcagaaaatacaaatgtaatttgcctttagcaccacacacacacacac 662
QY 659 atgcaaaaaccccgaggaacaaatttgtgtgtgcagatatgataaggaatgtg 718
DB 663 atacaaaagacacttaacaaactagattgttcccaactccctccttaagttatat 722
QY 729 ttggaaaaaagaattctgtatcagacagattctcagaacattcttaccctt 778
DB 733 ttggaaaaaagaattctgtatcagacagattctcagaacattcttaccctt 782
QY 779 gtggcagagtgatctcagatcatlgtagtaataatcatgttcttctgtgtgatatca 838
DB 783 gtcctcgagatgctgatacttcttctgtcagcaatgcttataatttggatgtg 842
QY 839 acacaaaataatgacactgagcagacagatgatatgtgtccacacactctgtgaa 898
DB 843 acagtaagaactttaacacagagctgtgtgtatctatcacaataatccagcagaa 902
QY 899 catctgtcacaataatcttaccagtgagcagcagatgtgtgtgtaattccgggat 958
DB 903 ctctgtcacaataatcttaccagtgagcagcagatgtgtgtgtaattccgggat 962
QY 959 ttgactggggagtgagacacaaaactctggaataatgcaacacacacactctglaag 1018
DB 963 atgactggggagtgagacacaaaactctggaataatgcaacacacacactctglaag 1022
QY 1019 acagagtgacagatataatgagtgccctacacagatgtgacagaggttcagactgctt 1078
DB 1023 acatgtgtacagcatgatttaccacatgtgtgaaagtggtgcaagacactgttgg 1082
QY 1079 caatccagagcgttgaaaatgtgtgtctgtgagtgagcaacactcattacataga 1138
DB 1083 ctgaccccaagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1142
QY 1139 atattctgaaatggctgacgtgattcatctgaggtgtgtgtgtgtgtgtgtgtgtgt 1198
DB 1143 agattccttttaacatcactgtgacttattctgtgcaatgagatgcccctcaagaagtt 1202
QY 1199 acaatgaataatccatctgat 1220
DB 1203 acaatgaataatcttctatgat 1224

```

```

RESULT 8
AAH57420
ID AAH57420 standard; cDNA; 1695 BP.
XX
AC AAH57420;
XX
DT 10-SEP-2001 (first entry)
XX
DE Human stomach cell specific cDNA sequence SEQ ID NO:260.
XX
KW Human; tissue specific; diagnosis; brain; heart; skeletal muscle;
KW lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;
KW metabolic disease; developmental disease; cytostatic; immunomodulatory;
KW neuroprotective; gene therapy; cancer; immunopathology; neuropathology.
XX
OS Homo sapiens.
XX
PN W0200132927-A2.
XX
PD 10-MAY-2001.
XX
PF 02-NOV-2000; 2000MO-US30396.
XX
PR 04-NOV-1999; 99US-0163508.
XX
PA (INCYTE GENOMICS INC.
XX
PI Sornasse T, Sellhammer JJ, Watson GA;
XX
DR WPI; 2001-291057/30.
XX
PT New cell and tissue specific polynucleotides useful for diagnosis,
PT prognosis or monitoring of treatments for disorders where the gene is
PT associated with a cancer, immunopathology or neuropathology -
XX
PS Claim 1; Page 189-190; 327pp; English.
XX
CC AAH57161 to AAH57576 represent cell and tissue specific polynucleotide
CC sequences (I). (I) can have cytosolic, immunomodulatory and
CC neuroprotective activities, and can be used in gene therapy. (I) and
CC proteins (II) encoded by them are used in high throughput screening
CC assays to select DNA molecules, RNA molecules, peptide nucleic acids,
CC mimetics, peptides, proteins, agonists, antagonists, antibodies or
CC their fragments, immunoglobulins, inhibitors, drug compounds and
CC pharmaceutical agents. Expression of (I) in a sample indicates the
CC differentiation of embryonic stem cells into a tissue selected from
CC brain, heart, kidney, liver, lung, skeletal muscle or pancreatic
CC tissues. (I) and (II) are used to produce an expression profile that
CC defines a metabolic or developmental process, treatment, condition,
CC disease or disorder. The gene profile can be used for diagnosis,
CC prognosis or monitoring of treatments and for investigating a
CC predisposition to a disorder where the gene is associated with a
CC cancer, immunopathology or neuropathology.
XX
SQ Sequence 1695 BP; 480 A; 356 C; 389 G; 470 T; 0 other;

Query Match 36.9%; Score 468.4; DB 22; Length 1695;
Best Local Similarity 64.1%; Pred. No. 1,8e-138;
Matches 706; Conservative 0; Mismatches 396; Indels 0; Gaps 0;
QY 119 ctgtgacccaagaagcatatgataatagtgaaatcatcacaatcaagctatccct 178
DB 116 ctggaagccctgaagtacatgaaacttagtcagatgattctatttggtatcccaa 175
QY 179 gtaagaataatgagtcgaactgaagaatggtatctcttctgttaagaagattccctc 238
DB 176 atgaagaataatgagtcgactgaagaatggtatctcttctgttaagaagattccctc 235
QY 239 gaggcctagtcgaacacagaggttcacagcctgtgtgttaactgacagatgccc 298
DB 236 atgggaagaataatcagaggaatacagcagagacccgtgtgtttttgacagctggt 295

```



[illegible]

16-DEC-1993:	53603-FR01260.
16-DEC-1992:	92FR-0015201.
(LJOU ) INST RECH JUVENILAL.	
Benicourt C, Blanchard C, Junien J;	
WPI, 1994-217890/26.	
P-PSDB; AAR56870.	
Recombinant canine gastric lipase and nucleic acid encoding it -	
are used for improving absorption of ingested fat, treating	
muoviscidosis etc. and in enzymatic bio-conversions	
Claim 1; Fig 8; 52pp; French.	
The sequence given below is the sequence of Figure 8, altered	
according to the amendments described on page 2 of the appended	
letter.	
CGI is used to improve absorption of ingested fat, in healthy and	
sick patients (e.g. having altered levels of gastric lipase); to	
treat conditions associated with insufficiency (or lack) of lipases,	
esp. mucoviscidosis or exocrine pancreatic insufficiency and partic.	
where immobilised, for bioconversions, e.g. hydrolysis or	
transesterification (other mammalian gastric lipases, or derivs.,	
can be used in this application).	
Sequence 1528 BP; 394 A; 386 C; 328 G; 420 T; 0 other:	
Query Match	35.0%; Score 444.2; DB 15; Length 1528;
Best Local Similarity	62.8%; Pred. No. 8.3e-131;
Matches 689; Conservative 0; Mismatches 408; Indels 0; Gaps	
125 acccggaagcattcttgaatattgaatgaatcacaatcacaagctatccctgtgagg	184
26 acccggaagtgaccatgataatgaatgaatgatactactacgtgggataaccagcttggg	85
135 aatatgaagtcgcaactggaagtgtgatalccttctgttaacaggaattcctcgaggcc	244
86 aatatgaagctgtgacgcaagagcggcttatatccttggatcgacagaattccttaaggga	145
245 tagtgaacctgaaggaagcaggttcccaagcctgtgtgttactgtgaatgagtcctagttg	304
146 ggaataatctcgagagataatagcgcggaagaccgttgcacatttggcaaacagcttgctcg	205
305 gaggtgcagcagcagcagattcccaacctgcgcacaaatagcctgtggtcttcatctggcag	364
206 catcgccacaacaactgcatctcccaactcctgcacaaacaaacagctgtgcttcatcctgcgcg	265
365 atgctgtgttttgacgtctgtgatactgggaaacagcaggggaaacgcgtgtgtctgaaacaca	424
266 acgcgcgggtagcagctgtgtgctggggaaacagcaggggcaaacctgggcaggaaggaatc	325
425 agaaccttccatagaccacaagatgaggttctgtggtcttcaagtatgatatgagtgcttagt	484
326 tgaactactgcgccgactccgtgcgaattctgtggtcttcaagcttggacgaagatggttaaat	385
485 ttgacctcttcggcagctgataaacttatttggacgaagaacggcgcaagaagaagactat	544
386 atgacctcttcggcagcacttgaattcatccttgaagaagaacggcgacagaagaagcact	445
545 atgtcgtctatccacagggcaccacacatggtgcttattgcatatttccacacatgcagagc	604
446 acgttgcatttcccaaggcaccacacattgttctcatcgtccttccacacacccaagc	505
605 tggctcagaataatcaatgtatttggctttagaccccatatggccactgtttaagctgaa	664
506 tggcgaagaacggaatacaaaccttcatgtatgaagctccgctgtgcacgcgtgaagtaacgcg	565
665 aaagcccccggagccaacattcttgtgcgcagatatagatcaaggaattgtttggca	724

Db 566 aaaccctgttaaaacacatcgtctgccttccttccttcacgttatattgaa 625  
 Qy 725 aaaaagaattctgtatcagacacagattctcgaacacatgttattacattgtgac 784  
 Db 626 acaaaattcttaccacacacacattcttgaataattctcgcacccagagatgtccc 685  
 Qy 785 aggtgattcttgatcagaattgttagtaataatcattccttcgttggtgattcaacca 844  
 Db 686 gcgagacggttgatctctctcgcagcaagccctgtttatcattgttgattgacacta 745  
 Qy 845 acaatataacatgagccgagcaagtgatattgtcgcacacacattcttgatgaacatctg 904  
 Db 746 tgaacttgaaacatgagcttgcttgatgtgattcgtcacaataccagcaggaacacatcg 805  
 Qy 905 tgcacaaatcttaccatcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 964  
 Db 806 ttccgaaacgtgtcctcactcgtgcacccggtgttgaatcgttggaagttccaagctttgact 865  
 Qy 965 ggggaggtgagacacaaatctggaataatgcaatcagcacaactcctgtaaggtacagag 1024  
 Db 866 ggggagccagcttcagacacatgacactatcagacagacagcagcagcagcagcagcagc 925  
 Qy 1025 tcagagataatgacggtccctcactcagacatgtgacagcagcagcagcagcagcagcagc 1084  
 Db 926 tgcagacacatgcatgtgcacacatgcagcagcagcagcagcagcagcagcagcagcagc 985  
 Qy 1085 cagaagacgtgaaatgtcgtctcctcgtgaggtgagcagcagcagcagcagcagcagcagc 1144  
 Db 986 ctcaagatgttgacactttgtcttcacaaagctcccaatctccttaccacaggaagatcc 1045  
 Qy 1145 ctgaatgggtcgaatgattcattcctggtggttgatgctcctcactacgtatgataatg 1204  
 Db 1046 ctcttcaatacactgacttacttctggtcagatgagcagcagcagcagcagcagcagcagc 1105  
 Qy 1205 aatcatccatcagtgatg 1221  
 Db 1106 aatgttccatgacg 1122

RESULT 11  
 AAT58915  
 ID AAT58915 standard; cDNA to mRNA; 1528 BP.  
 AC AAT58915;  
 AT 19-AUG-1997 (first entry)  
 DE Dog gastric lipase coding sequence.  
 KW Dog; duodenal; gastric; lipase; transgenic; plant; recombinant; extract;  
 KW food; absorption; fat; pancreatic; cystic fibrosis; exocrine; dairy;  
 KW hydrolysis; trans-esterification; substrate; enzyme; biofuel; ss;  
 OS Canis familiaris.  
 FH Key Location/Dualifiers  
 FT CDS 1..1140  
 FT /\*tag= a  
 FT /product= gastric lipase  
 PN MO9633277-A2.  
 PD 24-OCT-1996.  
 PE 19-APR-1996; 96MO-FR00606.  
 PR 20-APR-1995; 95FR-0004754.  
 PA (BIOC-) BIOEM SA.  
 PA (LJOU) INST RECH JOUVEINML.  
 PI Baudino S, Benicourt C, Cudrey C, Gruber V, Lenee P,  
 PI Merot B;

XX WP1: 1996-485783/48.  
 DR P-PSDB; AAM09382.  
 XX Recombinant human or canine pre-duodenal lipase proth. in transgenic  
 PT plants - useful for facilitating absorption of fat, as bio-catalysts  
 PT and for prodn. of bio-fuel  
 PS Claim 2; Fig 1; 130pp; French.  
 XX This is the nucleotide sequence encoding the dog pre-duodenal (i.e.  
 CC gastric) lipase enzyme. The sequence can be used to generate transgenic  
 CC plants producing recombinant lipase in an enzymatically active form.  
 CC Alternatively the first 12 bases of this sequence (encoding amino acids  
 CC 1-4 of the protein encoded by this sequence) or bases 1-162 of this  
 CC sequence (encoding amino acids 1-54) can be deleted to form the  
 CC derivatives designated delta-4 or delta-54 respectively. Plants, or  
 CC their extracts, expressing the lipases or the truncated derivatives, can  
 CC be used: (a) as pharmaceuticals or food to facilitate absorption of fat,  
 CC either in healthy subjects or in patients with inadequate levels of  
 CC gastric/pancreatic lipase e.g. those with cystic fibrosis or exocrine  
 CC pancreatic insufficiency, in the elderly or in patients undergoing medical  
 CC treatment which alters fat absorption; (b) for performing industrial or  
 CC agricultural reactions, e.g. in processing of fats or in the dairy  
 CC industry, for hydrolysis or transesterification reactions, etc., where  
 CC the plant material may provide both enzyme and substrate. The transgenic  
 CC plants can also be used for biofuel production.  
 SQ Sequence 1528 BP; 392 A; 387 C; 329 G; 420 T; 0 other;

Query Match 35.0%; Score 444.2; DB 17; Length 1528;  
 Best Local Similarity 62.8%; Pred. No. 8.3e-131;  
 Matches 689; Conservative 0; Mismatches 408; Indels 0; Gaps 0;

Qy 125 acccagaagcattcatgatatattgaatcattcaacatcagaagcattcctgtgag 184  
 Db 26 acccgaagtgcacatgatatataagcattgatacctactggtgacacagctgag 85  
 Qy 185 aatatgaagtcgacactgaagaatggtgatatctctcgttaacagagatccctcagcc 244  
 Db 86 aatatgaagttgtgacccgaagcaggttatatctcgttgatcagcagaattcctattgga 145  
 Qy 245 tagtgcacctaagaagaagcgttccagcctgtgtgttactacagcagatgctagt 304  
 Db 146 ggaataattcagagatatagccgagacgtgtgcatatttgacaacaggttgcg 205  
 Qy 305 gagggtgctagcactgatttccaactgcgcacaaatagcctgtggttcatctcggag 364  
 Db 206 catcgacacacaaactgcatctccaactgcgcacaaatagcctgtggttcatctcggag 265  
 Qy 365 atgctggttttagctgtgagatgggaacagaggggaacgcccgtgtcgaacacaa 424  
 Db 266 acgcccggatgaagcgtgtgtcgtgggaacagcaggggaacacccctggccaggaagatc 325  
 Qy 425 agacacttccatcagacaaagatggttctggtcttcaagtatgatagatgctagt 484  
 Db 326 tgtactactgcgcgcactcgcgcagcattcgtggtcttcaagcttgaagatggttaat 385  
 Qy 485 ttgaccttccgcaatgataaacttatttgcagaaaacggccaggaagaaatcttt 544  
 Db 386 atgaccttccgcaacttgcattcatttgaagaaaacgggaagacagacagctaacct 445  
 Qy 545 atgtcgtcattcagacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 604  
 Db 446 acgttgcattcaccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 505  
 Qy 605 tggctcagaataatcaaatgtatttgccttagcaccatagccatgcttaagatgaa 664  
 Db 506 tggcgaacagatcaaaaacttctatgattagctcccggttgcacagctgaagtagcacg 565  
 Qy 665 aaagcccgagcacaattttgtgtcgtgcagagatgatatgataagagatgttggca 724

Db 566 aaaccgttaacaacatcgtctgccttccttccttcacgttatatttgaa 625  
 QY 725 aaaaagaattctctatcagaccgattctcagacaactgttatttaacttctgccc 784  
 Db 626 acaaatattctaccacacactcttctgtatcaattctcgcacccagatgctccc 685  
 QY 785 aggtgattctgtacagatttctgtatcatatcatcttactctgtgtgtgttcaacaca 844  
 Db 666 gcgagacggtggtatctctctgcagcaacgcccgtttatcatattgttgatttgacata 745  
 QY 845 acaatatgaacatgagccagcaagtgtatagtctgcccacactcttctgtgaacatctg 904  
 Db 746 tgaacttgaacatgagctgctgtgtatctgtatctgtacataaaccagcaagaaatcgtg 805  
 QY 905 tgcataatctctacatctgagccagcagtgtaattctgtgtgaactccggcatttgact 964  
 Db 806 ttccagaacgtgtccacatgctccacagctgttaagtctgggaagttccaagcttttgact 865  
 QY 965 gggggagtgagacccaataatctggaataatgcaatcagccactctgttaaggtacagag 1024  
 Db 866 ggggaagccaggttcagacaatgacatcatcatcagagacatgctccctactacaacc 925  
 QY 1025 tcagagatatagcgttccctcacaagaatgtgacagaggtcagagactggttcaaatc 1084  
 Db 926 tgacagacatgcatggtgcacatcgcaggtgtgaaagcgtgcaacagacttctgtgcgac 985  
 QY 1085 cagaagacgtgaaatgtctctctgtgaggtgacacactatctacataagaatattc 1144  
 Db 986 ctccagatgttgaccttctgttcacagctcccaatctcatcttaccacaggaagattc 1045  
 QY 1145 ctgaatggtccagctgattctatctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1204  
 Db 1046 ctcccttaacaatcactgtgacttctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1105  
 QY 1205 aaatccatccatctgag 1221  
 Db 1106 aaatgttccatctgag 1122

RESULT 12  
 AA068389  
 ID AA068389 standard; DNA; 1531 BP.  
 AC  
 XX AA068389;  
 DT 20-FEB-1995 (first entry)  
 XX  
 DE Canine gastric lipase (Met-CGL).  
 XX  
 KW Canine gastric lipase; CGL; dog; fat; mucoviscidosis; enzyme;  
 KM bioconversion; exocrine pancreatic insufficiency; ss.  
 XX  
 OS Canis familiaris.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1143  
 FT /\*tag= a  
 XX  
 PN W09413816-A.  
 XX  
 PD 23-JUN-1994.  
 XX  
 PF 16-DEC-1993; 93MO-FR01260.  
 XX  
 PR 16-DEC-1992; 92FR-0015201.  
 XX  
 PA (LJOU ) INST RECH JOUVEINAL.  
 XX  
 PI Benicourt C, Blanchard C, Junien J;  
 XX  
 DR WPI; 1994-217890/26.  
 DR P-PSDB; AAR56871.  
 XX

PT Recombinant canine gastric lipase and nucleic acid encoding it -  
 PT are used for improving absorption of ingested fat, treating  
 PT mucoviscidosis etc. and in enzymatic bio-conversions  
 XX  
 PS Claim 2; Fig 8; 52pp; French.  
 CC  
 CC The sequence given below is the sequence of figure 8, altered  
 CC according to the amendments described on page 2 of the appended  
 CC letter.  
 CC CGL is used to improve absorption of ingested fat, in healthy and  
 CC sick patients (e.g. having altered levels of gastric lipase); to  
 CC treat conditions associated with insufficiency (or lack) of lipases,  
 CC esp. mucoviscidosis or exocrine pancreatic insufficiency and partic.  
 CC where immobilised, for bioconversions, e.g. hydrolysis or derivs.,  
 CC transesterification (other mammalian gastric lipases, or derivs.,  
 CC can be used in this application).  
 XX  
 S0 Sequence 1531 BP; 395 A; 386 C; 329 G; 421 T; 0 other;

Query Match 35.0%; Score 444.2; DB 15; Length 1531;  
 Best Local Similarity 62.8%; Pred. No. 8.3e-131;  
 Matches 689; Conservative 0; Mismatches 408; Indels 0; Gaps 0;

QY 125 acccagaagcttatgatattgtgaatcatccacatcaagcgtatccctgtgag 184  
 Db 125 acccagaagcttatgatattgtgaatcatccacatcaagcgtatccctgtgag 88  
 QY 29 accctgaagtgacatgatataagtcagatcatccatctgtggataccacagctgag 88  
 Db 29 accctgaagtgacatgatataagtcagatcatccatctgtggataccacagctgag 88  
 QY 185 aatltgaatgtgacacatgaatgtgatactcttctgttcaagaaggtatccagagcc 244  
 Db 185 aatltgaatgtgacacatgaatgtgatactcttctgttcaagaaggtatccagagcc 244  
 QY 89 aatatgaatgtgtgacagcaagcgttatcatccttggatctgacagaattccttatgga 148  
 Db 89 aatatgaatgtgtgacagcaagcgttatcatccttggatctgacagaattccttatgga 148  
 QY 245 tagtgcacctaagaagacaggttccagcctgtgtgtgttactgtcagacatgacctagt 304  
 Db 245 tagtgcacctaagaagacaggttccagcctgtgtgtgttactgtcagacatgacctagt 304  
 QY 149 ggaataatcagagatatagccggagacctgttgcatttttgcacaacagttgtctcg 208  
 Db 149 ggaataatcagagatatagccggagacctgttgcatttttgcacaacagttgtctcg 208  
 QY 305 gaagtgctagacaactgtgattccaactgcgcacacaatagcctgtgtctatcttcgag 364  
 Db 305 gaagtgctagacaactgtgattccaactgcgcacacaatagcctgtgtctatcttcgag 364  
 QY 209 catagccacaacatggtatctccaactgcgcacacaatagcctgtgtctatcttcgag 268  
 Db 209 catagccacaacatggtatctccaactgcgcacacaatagcctgtgtctatcttcgag 268  
 QY 365 atgtcgtgttggacgtgtgagtgaggaacagcaggggaacgcgtgtctgaaacaca 424  
 Db 365 atgtcgtgttggacgtgtgagtgaggaacagcaggggaacgcgtgtctgaaacaca 424  
 QY 269 acgcgggtgaagacgt 328  
 Db 269 acgcgggtgaagacgt 328  
 QY 425 agacactctccatagaccagaatgagttcgtgttccagtatgatatgagtgctaggt 484  
 Db 425 agacactctccatagaccagaatgagttcgtgttccagtatgatatgagtgctaggt 484  
 QY 329 tgtactactgcgccgactcgtcgaattctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 388  
 Db 329 tgtactactgcgccgactcgtcgaattctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 388  
 QY 485 ttgaccttccctgcagtgataacttattttgcaggaacagcggcaggaagaatctatt 544  
 Db 485 ttgaccttccctgcagtgataacttattttgcaggaacagcggcaggaagaatctatt 544  
 QY 389 atgaccttccctgcagtgataacttattttgcaggaacagcggcaggaagaatctatt 448  
 Db 389 atgaccttccctgcagtgataacttattttgcaggaacagcggcaggaagaatctatt 448  
 QY 545 atgtcgtgttccatcaagcagccacacatgagcttatttgacatttccacatccagag 604  
 Db 545 atgtcgtgttccatcaagcagccacacatgagcttatttgacatttccacatccagag 604  
 QY 449 acgttgcacattccacagggacacacatggtttcatctgtccttccacacatccagag 508  
 Db 449 acgttgcacattccacagggacacacatggtttcatctgtccttccacacatccagag 508  
 QY 605 tgcctcagaaatacaaatgatttctttagacacacatgacacgtgttaagatgaa 664  
 Db 605 tgcctcagaaatacaaatgatttctttagacacacatgacacgtgttaagatgaa 664  
 QY 509 tggcgaacagatcaaaacttctcatgcatgagctcgttgcacacggttagatcccg 568  
 Db 509 tggcgaacagatcaaaacttctcatgcatgagctcgttgcacacggttagatcccg 568  
 QY 665 aaagcccccggagccaataatttctgtgcagcagatagatgataagagatgtgtgca 724  
 Db 665 aaagcccccggagccaataatttctgtgcagcagatagatgataagagatgtgtgca 724  
 QY 569 aaaccctgttaacaacatcagctcgtccctctgtcccttcaagcttatatttgaa 628  
 Db 569 aaaccctgttaacaacatcagctcgtccctctgtcccttcaagcttatatttgaa 628  
 QY 725 aaaaagaattctgtatcagaccagattctcagacaactgttatttaacttctgtgccc 784  
 Db 725 aaaaagaattctgtatcagaccagattctcagacaactgttatttaacttctgtgccc 784  
 QY 629 acaaatattctaccacacacactcttctgtatcaattctcgcacccagatgctccc 688  
 Db 629 acaaatattctaccacacacacactcttctgtatcaattctcgcacccagatgctccc 688  
 QY 785 aggtgattctgtacagatttctgtatcatatcatcttctgtgtgtgtgtgtgtgtgtgt 844  
 Db 785 aggtgattctgtacagatttctgtatcatatcatcttctgtgtgtgtgtgtgtgtgtgt 844  
 QY 689 gcgagacggtggtatctctctgcagcaacgcccgtttatcatattgtgtgtgtgtgtgt 748  
 Db 689 gcgagacggtggtatctctctgcagcaacgcccgtttatcatattgtgtgtgtgtgtgt 748  
 QY 845 acaatatgaacatgagccagcagtgatattgtctgtccacacactctgtcgtgacatctg 904



```

XX OS Homo sapiens.
XX PN MO200102563-A2.
XX PD 11-JAN-2001.
XX PF 16-JUN-2000; 2000MO-JP03943.
XX PR 02-JUL-1999; 99JP-0188835.
XX PA (SAGA ) SAGAMI CHEM RES CENT.
XX PA (PROT-) PROTEGENE INC.
XX PI Kato S, Kimura T;
XX DR WPI: 2001-071581/08.
XX DR P-PSDB; AAB61608.
XX PT New human proteins with hydrophobic domains, useful for the treatment
XX PT of immune disorders, tumors, allergic conditions, thrombosis and
XX PT microbial infection -
XX PS Claim 4; Pages 122-124; 153pp; English.
XX CC The present invention relates to human proteins (AAB61608-AAB61617) and
XX CC their coding sequences (AAF28679-AAF28698). The proteins of the present
XX CC invention have hydrophobic domains and can be used for the treatment of
XX CC various immune deficiencies and disorders, such as severe combined
XX CC immunodeficiency (SCID), multiple sclerosis, rheumatoid arthritis,
XX CC autoimmune pulmonary inflammation, graft-versus-host disease and
XX CC Guillain-Barre syndrome. The proteins may also be useful in the treatment
XX CC of allergic reactions and conditions, such as asthma and in regulation of
XX CC haematopoiesis or lymphoid cell deficiencies. The proteins may also have
XX CC utility in compositions used for bone, cartilage, tendon and/or nerve
XX CC tissue growth or regeneration as well as wound healing and in the
XX CC treatment of burns. The proteins may be used in the treatment of
XX CC periodontal disease and in other tooth repair processes. Other uses
XX CC include treatment of thrombolytic and haemostatic conditions, treatment
XX CC or prevention of tumours and inhibiting infection by bacteria, viruses,
XX CC fungi and other parasites.
XX SQ Sequence 1308 BP; 402 A; 246 C; 264 G; 396 T; 0 other;

Query Match      22.5%; Score 286; DB 22; Length 1308;
Best Local Similarity 67.4%; Pred. No. 1.8e-80;
Matches 403; Conservative 0; Mismatches 195; Indels 0; Gaps 0;

```

```

OY 446 atgagttcgtggttcacgtatgatgatgagttgagttgaccttcctcagtgataa 505
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 479 cagataactggcgcttcagttgagtgatgagtcgaataatgaccttcacgacaaatca 538
OY 506 acttattttcagaagaaacgggccaagaaagatctatagtctggtcattcaagagca 565
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 539 atttataatagaaaactggacagaagcgacttactatagctggtccactcaaaagca 598
OY 566 ccacatgggcttattcattcttcacacatgcagagctggtccagaagaatcaaatgt 625
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 599 ccacatagcttattatagcatttctctacaacccagaactgtgctcaaaagttatagat 658
OY 626 atttgcttagcaccacatagcaccactgttgaagcatgcaaaaagcccgagccaatt 683
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 659 ttttgacatgctccagcttgcacaggttaataatcaccccaagtcctatgaaanaact 716

RESULT 15
AAA44349
ID AAA44349 standard; CDNA; 617 BP.
XX AC AAA44349;
XX DF 21-AUG-2000 (first entry)
XX DE Human secreted expressed sequence tag SEQ ID NO:924.
XX KW Human; mouse; chicken; rat; secreted expressed sequence tag; SEST;
KW expressed sequence tag; EST; probe; chemotactic; proliferative;
KW immunomodulatory; haematopoietic; chemokine; analgesic; haemostatic;
KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW antiviral; antidiabetic; antiasthmatic; vulnery; antiparkinsonian;
KW antitumor; osteoprotective; neuroprotective; neotrophic; antipsoriatic;
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW central nervous system disorder; Alzheimer's disease; stroke;
KW Parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW tumour; infection; depression; psoriasis; ss.
XX OS Homo sapiens.
XX OS MO200021991-A1.
XX PN 20-APR-2000.
XX PD 15-OCT-1999; 99WO-US24206.
XX PF 15-OCT-1998; 98US-0104436.
XX PR 15-OCT-1998; 98US-0104436.
XX PA (GENMY ) GENETICS INST INC.
XX PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
XX PI Merberg D, Treacy M, Bowman MR;
XX DR WPI: 2000-317938/27.
XX XX
XX PT Isolated polynucleotides, and encoded proteins, comprising secreted
XX PT expressed sequence tags (SESTs), useful for treating various disorders -
XX PT such as autoimmune, infections, and central nervous system disorders -
XX PS Claim 1; Page 441; 803pp; English.
XX CC AAA44342 to AAA45925 represent specifically claimed secreted expressed
XX CC sequence tags (SESTs), isolated from human, mouse, chicken and rat
XX CC tissue sources. The SESTs can have a range of activities depending on
XX CC the tissues they were isolated from. The activities include:
XX CC chemotactic; proliferative; immunomodulatory; haematopoietic;
XX CC chemokine; analgesic; haemostatic; thrombolytic; antiinflammatory;
XX CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
XX CC antiasthmatic; vulnery; antitumor; osteopathic; neuroprotective;

```



CC neurotrophic; antiparkinsonian; antipsoriatic; cerebroprotective;  
CC anticonvulsant; and antidepressant. The SESSs can be used for gene  
CC therapy and in vaccines. The SESSs are useful as probes for the  
CC identification and isolation of full-length cDNAs and genomic DNA  
CC molecules which correspond to the SESSs. Proteins encoded by the SESSs  
CC are useful in assays for determining biological activity and raising  
CC antibodies. They may be useful for treatment of autoimmune disorders  
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions  
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,  
CC osteoporosis, osteoarthritis, central nervous system disorders  
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation  
CC disorders (hemophilia, thrombosis), inflammatory disorders (Crohn's  
CC disease), tumours, bacterial, fungal or viral infections, depression and  
CC psoriasis. AAA45526 to AAA55931 represent linker variants which are given  
CC in the exemplification of the present invention.

**SQ** Sequence 617 BP; 181 A; 125 C; 139 G; 172 T; 0 other;

Query Match	20.3%	Score 257.8	DB 21	Length 617
Best Local Similarity	70.7%	Pred. No. 1.1e-71		
Matches 343	Conservative 0	Mismatches 142	Indels 0	Gaps 0

QY	119	ctcggagcccggaagcattctatgatatctgtaacatcatccaacatcaagctatccct	178
Db	133	ctcgaaagcccggaagctgactatgtaaacctctgcgcgaatgatacttaattggtgataccaa	192
QY	179	gtcgaggaataatgaagctgcgaactgaagaatggtataltcccttcgttaaacaggaattcc	238
Db	193	atgaaagaaataatgaagctgtgactgtaagaatggtataltcttgaaagtcaataagaattcc	252
QY	239	gaagcctagctgcaaaccttaagaagaacagaggttcagagcctctggtgttactgcaagatg	298
Db	253	atgggaagaanaaatctcagaggaaataacagccagagagacctgtgtgttttcgacatggt	312
QY	299	tagttgtgaagtgtctagcaactgtgaattccaaactgtgccaaacaatagcctggtctta	358
Db	313	tgccttgatactcgccacaacaactggaattctccaacctgcgcgaacaacagcctgtcttc	372
QY	359	tgggaagctgcggttttgagctgtagtggaggaaacagcaagggaagaagcctgtctcgaa	418
Db	373	tgggaagatgcggtttaaagatgtgtgtgctggcaacagcaagaaagaacacctgtgcca	432
QY	419	aaccagaagacactccatagacagcaacaaatagttctgtgcttcgaatgataagatgg	478
Db	433	gaaacctgttaactatctaccagatccagttgaattctgtgcttcgaatgataagatgg	492
QY	479	ctaggtttgaacctctgcgcgagtgataacttattttgcgcaaaaacggtgccaagaaaga	538
Db	493	ctaataatgaccttcagcccaactgacttaactatgttaaaagaaaacctggtgcagaagagc	552
QY	539	tctattatgctgctatcatcaagggcacaacatbggcttattgtcaatttccacaatgc	598
Db	553	tacctaatagtgtgcattccaccagggcacaccacatgttttattgtcctttccacaatc	612
QY	599	caagag 603	
Db	613	tcgag 617	

Search completed: January 30, 2002, 22:13:51  
Job time: 3898 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2002, 20:27:03 ; Search time 90.11 Seconds

(without alignments)  
3189,442 Million cell updates/sec

Title: US-09-578-063-46

Perfect score: 1269

Sequence: 1 atgttggaacctgtcgaag.....gacggtgtgagccgcatgtt 1269

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 segs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/2/1na/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/2/1na/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/2/1na/PCTUS.COMB.seq: \*  
6: /cgn2\_6/ptodata/2/1na/Backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	444.2	35.0	1137	1	US-08-227-108-2
2	444.2	35.0	1137	2	US-09-073-674-2
3	444.2	35.0	1140	1	US-08-227-108-4
4	444.2	35.0	1140	2	US-09-073-674-4
5	444.2	35.0	1146	1	US-08-227-108-6
6	444.2	35.0	1146	2	US-09-073-674-6
7	444.2	35.0	1528	1	US-08-227-108-1
8	444.2	35.0	1528	2	US-09-073-674-1
9	48.2	3.8	178	1	US-08-751-782-2
10	48.2	3.8	178	2	US-08-925-171-2
11	46.4	3.7	43676	3	US-09-356-952-12
12	36.6	2.9	3271	2	US-08-852-806-1
13	36.6	2.9	3271	3	US-09-163-669-1
14	36.6	2.9	3282	1	US-08-276-852-154
15	36.6	2.9	3282	1	US-08-276-852-154
16	36.6	2.9	3282	1	US-08-899-575-154
17	36.6	2.9	3282	1	US-08-899-575-154
18	36.6	2.9	3282	1	US-08-899-575-154
19	36.6	2.9	3282	1	US-08-899-575-154
20	36.6	2.9	3282	5	PCT-US95-08743-154
21	36.6	2.9	3282	5	PCT-US95-08743-154
22	36.6	2.9	13254	1	US-08-276-852-156
23	36.6	2.9	13254	1	US-08-276-852-156
24	36.6	2.9	13254	1	US-08-899-575-156
25	36.6	2.9	13254	1	US-08-899-575-156
26	36.6	2.9	13254	1	US-08-899-575-156
27	36.6	2.9	13254	1	US-08-899-575-170

28	36.6	2.9	13254	5	PCT-US95-08743-156	Sequence 156, App
29	36.6	2.9	13254	5	PCT-US95-08743-170	Sequence 170, App
30	33.6	2.6	8392	1	US-08-080-255-6	Sequence 6, Appl
31	33.6	2.6	8392	3	US-08-465-713-6	Sequence 6, Appl
32	33.6	2.6	8392	5	PCT-US93-05857-6	Sequence 6, Appl
33	32.6	2.6	7218	1	US-08-232-463-14	Sequence 14, Appl
34	32.6	2.5	2197	1	US-08-233-005-3	Sequence 3, Appl
35	32.6	2.5	2197	1	US-08-428-943-3	Sequence 3, Appl
36	32.6	2.5	2197	3	US-09-016-649-3	Sequence 3, Appl
37	32.6	2.5	2197	3	PCT-US95-04858-3	Sequence 3, Appl
38	32.6	2.5	8342	5	US-08-545-860D-63	Sequence 63, Appl
39	32.6	2.5	8342	5	PCT-US94-04496-63	Sequence 63, Appl
40	31.6	2.5	3343	4	US-08-965-762-19	Sequence 19, Appl
41	31.6	2.4	3212	3	US-08-673-814-1	Sequence 1, Appl
42	30.4	2.4	1529	1	US-08-336-778-1	Sequence 1, Appl
43	30.2	2.4	4982	3	US-08-699-103B-1	Sequence 1, Appl
44	30.2	2.4	10079	2	US-08-476-866-20	Sequence 20, Appl
45	30.2	2.4	3639	2	US-08-737-524B-26	Sequence 26, Appl

## ALIGNMENTS

```

RESULT 1
US-08-227-108-2
; Sequence 2, Application US/08227108
; Patent No. 5807726
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.75
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/227,108
; FILING DATE: 03-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci, Allan A.
; REGISTRATION NUMBER: 30,256
; REFERENCE/DOCKET NUMBER: 7620-033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8664/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1137 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1137
; US-08-227-108-2
;
; Query Match 35.0%; Score 444.2; DB 1; Length 1137;
; Best Local Similarity 62.8%; Pred. No. 1.5e-134;
; Matches 689; Conservative 0; Mismatches 408; Indels 0;

```

QY	125	accacgaagcatcattgaatcattagtgaaatacacaacaaacgaagcttcccttgag	184
Db	26	ACCTGGAAGTGACCATGATATATATAGTCAGATGATCACTTACTGGGGATACCCAGCTGAG	85
QY	185	aataatgaagtcgaactgtaagatgaggtatatacccttcctgttlaacagatctccctgagcc	244
Db	86	AATATGAAGTTGTGACGAGAGCGGTTATATCTTGGATGCGACGAATTTCTTATGGGA	145
QY	245	taatgcaacttaagaagacaggttccagagccgtgtgttactgtcagcatgtgctagtgtg	304
Db	146	GGAAAAATTTAGGAAATATAGGCCGGAGACCTGTTGCATTTTGTCAACACGGTTTGCTCG	205
QY	305	gaggtgtcgaactgtgatttccaaactccgccaacataagccctgagatcttcttcggag	364
Db	206	CATGACGCAAAACTGGATGATCCAACTCCGCCAACACAGCTGCGCTTACTCTGGCCG	265
QY	365	atgtcgtgtttgaagctgtgtgatagtgggaacagcagggaaacgcgtgtgttcgaaaacca	424
Db	266	ACGCGCGGATGACGAGTGTGTGCTGGGGAACAGCGGGGGCAACACCTGGCGCAGAGGAATC	325
QY	425	agaaacctccataagaacccaagatgagttctgagcttcaagttatgatgagatgagcttagt	484
Db	326	TGTACTACTGCCCGACTCCGTGCAATTTCTGGGCTTTACGCTTTGACGAGATGGCTTAAT	385
QY	485	ttgaaccttccgtcagtgatataaacttatttttcgagaaaacggcgccaggaagaatctat	544
Db	386	ATGACCTTCCCGCACCATTTGACTTCTATGTGAAGAAAAGGGACAGGACGATCTACT	445
QY	545	atgtcgtatctcaacaggacacacaagggtcttatgtcatlttccacatgacagagc	604
Db	446	ACGTTGGCCATTTCCACGGGGCACCCATGTGGTTACTCCCTTTTCCACCAATCCCAAGC	505
QY	605	tggctcgaataatcaaaaatgatttctgttgtagccacaaagcaactgttaagcgtga	664
Db	506	TGGCGAAACGGATCAAAACCTTTATGCAATTACTCCCTGCGACCGGAAGATCACCAGC	565
QY	665	aaagcccgagaccaaattttgtgtcgtcgagataatgaatgaagaggtatgtttgcga	724
Db	566	AAACCCGTGTTAAACAAACTGATCTGTCGCCCTTCGTTCCCTTCAAGCTTATATTGGAA	625
QY	725	aaaaagatttctgtatcagacagagttctcagaagaactgtttatattacccttggcc	784
Db	626	ACAAATATTTCTTACCCACACACACTTCTTTGATCAATTTTCCGCCACCGAGATATCTCC	685
QY	785	aggtgatcttgcatacagatttgcagtaatacatgtaacttcttggttgattcaacaca	844
Db	686	GCAGACGGTGGATCTCTCTGTGAGAACGCCCTGTTATCTTGTGTGATTTGACACTA	745
QY	845	acaaatgaacatgagccgagcaagtatatagtctgtgcacaaccttgcctgtgaacatctg	904
Db	746	TGAACCTTGAACATGAAATCGCTTGAGATGTATATCTGTCAATATATCAGAGGACATCGG	805
QY	905	tgcaaaatctcaacactgagacagcagtgtaattcttctgtgaacctccggagatcttagct	964
Db	806	TTTCAGAACGTGCTCCACTGTGTCACGAGCTGTAAAGTCTGTGGAAGTTCCAAAGCTTTTGACT	865
QY	965	gggggagtgtagaccaaanaactctggaanaaaatgcaatcagcacaactcctgtlaagtacagag	1024
Db	866	GGGGAAGCCAGTTTCAGAAACATGATSCATATATCATCAGACGTGCTCCCTACTACACC	925
QY	1025	tcagaaatagaaaggtccctcacaagaaatgtgacagggaggttaagacgtgcttcaaatc	1084
Db	926	TGACAGCATGATGTGTGCCAATGCGAGTGTGGAACGTGTGCAACACTTCTGTGGCCGACC	985
QY	1085	cagaagaagtgtaanaatgctgtctctctcgtgaggtgacaaactcactacataagaatatct	1144
Db	986	CTACAGCTGTGTGACCTTTTGTGCTTCCAACTCCCAATCTCATTTTACCAAGGAAGATTTC	1045
QY	1145	ctgaagaaggtcagcgagattcatctgtgtgtgttgatgtcctcactacagatgtacaagt	1204
Db	1046	CTCTTTCATCAATCTTGGACCTTATATGTGGCCATGAGATCCCTCTTAAACGGGTTTACATG	1105

Qy	1205	aatcatcctcatgtatg	1221
Db	1106	AAATGCTTCATGATG	1122

RESULT 2  
US-09-073-674-2

; Sequence 2, Application US/09073674  
Patent No. 5008100

; GENERAL INFORMATION:

```
;; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
```

APPLICANT: Junjen, Jean-Louis  
TITLE OF INVENTION: Recombinant Dog Gastric Lipase

NUMBER OF SEQUENCES: 21

;  
CORRESPONDENCE ADDRESS:  
;  
ADDRESSEE: Warner-Lambert Company

STREET: 2800 Plymouth Road  
CITY: Ann Arbor

STATE: Michigan  
COUNTRY: U S A

ZIP: 48105

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

;
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
;

```

```
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
```

CONSENT AFFILIATION DATA:  
APPLICATION NUMBER: US/09/073,674

```

; FILING DATE:
; CLASSIFICATION:
;

```

ATTORNEY/AGENT INFORMATION:  
NAME: Crissey, Todd M.

REGISTRATION NUMBER: 37,807

; TELECOMMUNICATION INFORMATION:

TELEPHONE: 734 622-75  
TELEFAX: 734 622-1553

TELEX: INFORMATION FOR SEO ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1137 base pairs  
TYPE: nucleic acid

```

; STRANDEDNESS: double
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: CDNA to mRNA
; FEATURE:

```

NAME/KEY:	CDS
1000000000	11111

LOCATION  
US-09-073-674-2

Query Match	35.0%;	Score 444.2;	DB 2;	Length 1137;
Best Local Similarity	62.8%;	Pred. No. 1.5e-134;		

Matches 689; Conservative 0; Mismatches 408; Indels 0; Gaps 0;

Oy 125 acccagaagcattcatgaatattagtgaaatcatccacatcaagcgtatccctgtagg 184

Db 26 ACCCTGAGTGACCATGATATAGTCAGATGATCACCTACTGGGGATACCCAGCTGAG 85

185 **ataatgaactccgaactgaagatgaatataatcctttctctatcaacaqaattcctcaagacc** 244

[illegible]

00 AHAIAWAAUIGIWAACGAWAALUGUIAIAIAICCIUGGWAICWAUAIIICCIHAIUGWA 143

001

QY 245 tagtgcacctagaagaacagttccacgcccgtgtgttactgcagcatgacctagtctg 304

Db 146 GGAATAATTCAGAGAATATAGGCCGGAGACCCTGTTCATTTTTCACAACACCGGTTGCTCG 205

Oy 305 gaagtgtacgaactggaatttccaactgcccacaacatagcctggtgttcattctgacag 364

Db 206 CATCAGCCACAACCTGGATCTCCAACTGCCCACAACAGCCCTGGCCCTTCATCTCGGCCG 265

365 atgcctggttttgacgtgtgatgggaacacagcaggggaacacgccttggtctcgaaacaca 424

```

Db 266 ACGCGGCTAGACCTGTGGCTGGGAACAGAGGGGCAACACCTGGGGCCAGAGAGATC 325
Qy 425 agacactctcatagaccagaatgagttctgggttcctcaatgataatgagatgagtgctaggt 484
Db 326 TGTACTACTGCGCGGACCTCGCTGCAATCTGGGGCTTTGACGCTTTCAGAGATGGCTAAT 385
Qy 485 ttgactctccgcagtgatgataacttatttgcagaaaagggccaggaagaaagactatt 544
Db 386 ATGACCTTCCGCGCACCACTTGAATCTTGAAGAAAAGGAGAGAGACGACGATCACT 445
Qy 545 atgtcgattctacagagggccaccacagggcttattgcaattcccccacatgacagagc 604
Db 446 ACGTTGGGCATTCACAGGGGCAACACATTTGTTTATGCGCTTTCCACACATCCCAAGC 505
Qy 605 tggctcagaanaatcaaatgtatatttgcttaagcaccatagccactgttaagcatgaa 664
Db 506 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCGTTGCCACCGTGAAGTACACG 565
Qy 665 aaagcccccggagaccgaattttgtgtcgcagagataatgataaggaattggttgca 724
Db 566 AAACCTGTAAACAACATGCTGTCCTTCCTTCCTTCAGCTTATATTGGAA 625
Qy 725 aaaaagaattctgtatcagaccagattctcagaacactgttattacatttggtgac 784
Db 626 ACAAATATTTTACACCAACACACTTCTTGAATCAATTTCTGCCACGAGATATGCTGCC 685
Qy 785 aggtatctctgatacagaattgttagtaataatcattctctgggtgagatcaaacaca 844
Db 686 GCGAAGCGGTGATCTCTCTGACCAAGCCCTGTTTATCATTTGTGATTTGACACTA 745
Qy 845 acaatatgaatagccagagcaagtgtatagtcgtccacactcttgctggaacatctg 904
Db 746 TGAACCTTAACATGATGATGCTTGTGATCTGTACATATCCAGAGAGAACATCGG 805
Qy 905 tgcgaatattctacacagggagccagggagtgatcttcgtgtaactccggagatttgact 964
Db 806 TTCAAGAGCTGCTCCACATGCTGTCACAGGCTGTTAAGCTGCGGAAGTTCCAACTTTTGACT 865
Qy 965 gggggagtgagaccgaataatctggaataatgcaatcagcaactccctgtaagtagcagag 1024
Db 866 GGGGAGGCCAGTTGAGAACATGATGACATATCATCAGACGATGCTCCCTACTACAAAC 925
Qy 1025 tcaagagatagacggttcccttaagcaatgtagcagagaggttaagagcttctcaatc 1084
Db 926 TGACGACATGATGATGCAATGCGAGGTGGAGACGATGCAACGACTTGTGCGGACG 985
Qy 1085 cagaagacgtgaaatgctgtctctgaggtgagacacactatctacataagaatatctc 1144
Db 986 CTCACGATGTGACCTTTTGTCTTCCAAAGCTCCCAATCTCATTTACCAAGAGAGATTG 1045
Qy 1145 ctgaatgggtcagtgatattcactgaggttgaggttgatgctccacatgataatg 1204
Db 1046 CTCCTTAAATCAATCACTTATCTGAGGCGCATGATGATGCCCTCAAGCGGTTTACATG 1105
Qy 1205 aatcatcatctgagt 1221
Db 1106 AAATTGTTCCATGATG 1122

```

## RESULT 3

US-08-227-108-4

Sequence 4, Application US/08227108

Patent No. 5807726

GENERAL INFORMATION:

APPLICANT: Blanchard, Claire

APPLICANT: Benicourt, Claude

APPLICANT: Junien, Jean-Louis

TITLE OF INVENTION: Recombinant Dog Gastric Lipase

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie &amp; Edmonds

STREET: 1155 Avenue of the Americas

```

CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/227,108
FILING DATE: 03-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fannuci, Allan A.
REGISTRATION NUMBER: 30,256
REFERENCE/DOCKET NUMBER: 7620-033
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1140 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1140
US-08-227-108-4

Query Match          35.0%; Score 444.2; DB 1; Length 1140;
Best Local Similarity 62.8%; Pred. No. 1.5e-134;
Matches 689; Conservative 0; Mismatches 408; Indels 0; Gaps 0;

Qy 125 accagaagcattcatgataatagtgataatccaaatcaagagctatccctgtgagg 184
Db 29 ACCCTGAAGTGCATGATATTAATCAATGATCACTACTGGGATACCACTGAGG 88
Qy 185 aatatgaagtcgacctaagaatggtatattccttctgttaacagagatccctgagggc 244
Db 89 AATATGAAGTTGTGACCCAGACGCTTATATCTTTGGATGACAGAAATTCCTATGGA 148
Qy 245 tagtcaacctaaagaacaggttccaggcctgtgtgttactgcaagatgctagttg 304
Db 149 GGAATAATTCAAGAAATATAGGCCGAGACCTGTTGCAATTTTGGCAACAGGTTGCTCG 208
Qy 305 gagggtgtagcaacttgatcttccaaactgcccacaataagcctgggtctcaattctggcag 364
Db 209 CATCGGCCCAAAACGTGATCTCCAACTGCCCAACACAGCCTGGCTTCATCTCGGCGG 268
Qy 365 atgtcgttttgagtggtgagtgaggaacagagggggaacgacctggtctcgaaacaca 424
Db 269 ACGCGGGGTACGACGTGTGGCTGGGGAACAGAGGGGCAACACCTCGGGCCAGAGATTC 328
Qy 425 agacactctcatagaccagaatgagttctgggttcctcaatgataatgagatgagtgctaggt 484
Db 329 TGTACTACTGCGCGGACCTCGCTGCAATCTGGGGCTTTGACGCTTTCAGAGATGGCTAAT 388
Qy 485 ttgactctccgcagtgatgataacttatttgcagaaaagggccaggaagaaagactatt 544
Db 389 ATGACCTTCCGCGCACCACTTGAATCTTGAAGAAAAGGAGAGAGACGACGATCACT 448
Qy 545 atgtcgattctacagagggccaccacagggcttattgcaattcccccacatgacagagc 604
Db 449 ACGTTGGGCATTCACAGGGGCAACACATTTGTTTATGCGCTTTCCACACATCCCAAGC 508
Qy 605 tggctcagaanaatcaaatgtatatttgcttaagcaccatagccactgttaagcatgaa 664
Db 509 TGGCGAAACGGATCAAAACCTTCTATGCATTAGCTCCGTTGCCACCGTGAAGTACACG 568

```



Db 809 TTCAAGAGTCTCTCACTAGTGGCCAGGCTGTAACTCTGGGAAGTTCACACTTTTACT 868  
Qy 965 ggggagtgagacaaaatctggaaaaatgcaatcagccactctgttaagatcagag 1024  
Db 869 GGGGAAGCCCGTTCAGAAATGATGACATCATCATGAGAGATGCGTCCCTACATCAAC 928  
Qy 1025 tcagagatagagcgtccctacagcaatgtagacagagagtcagagcttccaaac 1084  
Db 929 TGACAGCATGTCATGTCGCAATCCAGTGTGGAACGGTGCAACGACTTGTGCGCAC 988  
Qy 1085 cagaagacgtgaaaatgctgtctctgagtgagacaaacctatcacaataagatatc 1144  
Db 989 CTCACGATGTTGACCTTTTGTCTTCCAGCTCCCAATCTCATTTACACAGAGATTC 1048  
Qy 1145 ctgaatggctcagtgattcattcactcggggttggatgctccctaccggtatgacaatg 1204  
Db 1049 CTCCTTCAATCACTTGTGCTTATCTGGGCCATGATGATGCCCTCAACGGGTTTACATG 1108  
Qy 1205 aaatcattcatgctg 1221  
Db 1109 AAATGTTTCATGATG 1125

RESULT 5  
US-08-227-108-6  
; Sequence 6, Application US/08227108  
; Patent No. 5807726  
; GENERAL INFORMATION:  
; APPLICANT: Blanchard, Claire  
; APPLICANT: Benicourt, Claude  
; APPLICANT: Junien, Jean-Louis  
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Penlie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/227,108  
; FILING DATE: 03-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Faucel, Allan A.  
; REGISTRATION NUMBER: 30,256  
; REFERENCE/DOCKET NUMBER: 7620-033  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 790-9090  
; TELEFAX: 212 869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1146 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; US-08-227-108-6

Query Match 35.0%; Score 444.2; DB 1; Length 1146;  
Best Local Similarity 62.8%; Pred. No. 1.5e-134;  
Matches 689; Conservative 0; Mismatches 408; Indels 0; Gaps 0;

Qy 125 accaagaagcatcatgatatagtgaaatcatccacaatcaaggtatccctgtgag 184

Db 26 ACCCTGAAGTACCATGATATATAGTCAATGATGATCACCCTACTGGGATATACCGAGCTAGG 85  
Qy 185 aataagaagtcgaactggaagatggatatacctcttgttaacaggaattcctcaggcc 244  
Db 86 AATATGAAGTGTGATACCGAAGAGCGTTATATCCCTTGGAGATGACAGATTCCTATGGA 145  
Qy 245 taatgcaactaagaagaacaggtccagcctgtgtgttaactcgaacatgagcctatgt 304  
Db 146 GGAATAATTCAGAAATAATAGCGGAGACCTGTGATTTTTCACACAGGTTTGTCTG 205  
Qy 305 gaagtgcagaagaattgatttccaaactgccaacaaatagcctggatcattcttcag 364  
Db 206 CATAGCCACAAATGATCTCAACTGCCCCAACACAGCTGCTTCACTCTGAGCCG 265  
Qy 365 atgtcgttttgaagctgtgagtgaggacagcaggggaaacgcttgcttcgaacaaca 424  
Db 266 ACGCGGGGTACGAGCTGTGGCTGGGGAACAGCAGGGGCAACACTGGGCCAGAGATTC 325  
Qy 425 agaacactccatagacccaagaatgagttcgtggcttcagttatgatgatgagtgatg 484  
Db 326 TGTACTACTGCGCCGACTCCGTCGAATTCGGCTTTCAGCTTTCAGAGATGCTTAAT 385  
Qy 485 ttgaccttctgcagtgataaacttatttgcaagaaacgggacaggaagaatcatt 544  
Db 386 ATGACCTTCCCGCCACCATGACTTCATCTTGAAGAAACGGACAGACAGCTTACACT 445  
Qy 545 atgtcgcattcacaagggacacacacatgggcttatttgcatcttccacatgcagagc 604  
Db 446 ACGTTGGCCATTCCCGAGGACACACATGTTGTTTCACTGCGCTTTCACCAATGCCAAC 505  
Qy 605 tggctcgaagaaatcaaatgtatlttgcttagcaacatagccatgtaagcaatgaa 664  
Db 506 TGGCGAAACGGATCAAAACCTTATGATGATGATGATGATGATGATGATGATGATG 565  
Qy 665 aaagcccccggacaaatttltgtgtccagatagatgataagagatgtgtgtgca 724  
Db 566 AAACCTGTTAAACAACTCATGCTCGCTTGTGCTTCTTCACTTATATATGAA 625  
Qy 725 aaaaagaattctglatcagaacagatttccagacaactgttatttaccctgtgtg 784  
Db 626 ACAAATATTTCTACCCACACACACTTCTTGTGATCAATTTCTGCGCAGAGATGCTGCC 685  
Qy 785 aggtgattcttgatcagattttagtaatatcatgtaacttctgtgtgtgtgtgtgtgt 844  
Db 686 GCGAGACGCTGATCTCTCTGCAAGCAAGCCCTGTTATCATTTGTGATTTGACACTA 745  
Qy 845 acaatatgaatagagcagcaagtgatatagtctgccaacactctgtgtgtgtgtgtgt 904  
Db 746 TGAACCTGACATGATGCTTGGATGATGATGATGATGATGATGATGATGATGATG 805  
Qy 905 tgcacaatatctcaactcagtcagcagcagtgaaatctgtgtgaactccgggcatltgact 964  
Db 806 TTCAAGAGTCTCTCACAGTCCAGGCTGTAACTGGGAATTCACAGCTTGTGACT 865  
Qy 965 ggggagtgagacaaaatctggaaaaatgcaatcagcgaactcctgtlaagtgagag 1024  
Db 866 GGGGAAGCCCGTTCAGAAATGATGATGATGATGATGATGATGATGATGATGATG 925  
Qy 1025 tcagagatagagcgtccctacagcaatgtagacagagagtcagagatgcttcaaatc 1084  
Db 926 TGACAGATGTCATGTCGCAATCCAGTGTGGAACGGTGGAACGACTTGTGCGCAC 985  
Qy 1085 cagaagacgtgaaaatgctgtctctgagtgagacaaacctatcacaataagatatc 1144  
Db 986 CTCACGATGTTGACCTTTTGTCTTCCAGCTCCCAATCTCATTTACACAGAGATTC 1045  
Qy 1145 ctgaatggctcagtgattcattcactcggggttggatgctccctaccggtatgacaatg 1204  
Db 1046 CTCCTTCAATCACTTGTGCTTATCTGGGCCATGATGATGCCCTCAACGGGTTTACATG 1105  
Qy 1205 aaatcattcatgctg 1221



MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/227,108  
FILING DATE: 03-APR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fanucci, Allan A.  
REGISTRATION NUMBER: 30,256  
REFERENCE/DOCKET NUMBER: 7620-033  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1528 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA to mRNA  
US-08-227-108-1

Query Match 35.0%; Score 444.2; DB 1; Length 1528;  
Best Local Similarity 62.8%; Pred. No. 1.8e-134;  
Matches 689; Conservative 0; Mismatches 408; Indels 0; Gaps 0;

QY 125 acccagaacatcagaatattagtaaatcacaatcaagaactccctgtgag 184  
DB ACCCTGAAGTGCATGATATTAAGTCAATGATCAGTACCTACTGGGATACCCAGCTGAG 85

QY 185 aataagaatgcgaactgaagaatggtatattccttcgttaacagagatcctcgaagcc 244  
DB AATAAGAAGTGTGACCCAGAGCGGTATATCTTGGGATCGACAGAAATTCCTTTGGGA 145

QY 245 tagtgaacccaagaagaagatcagtcagccctgtgttactgaagcctgagctgtg 304  
DB GGAAGAAATTCAGAGATATAGGCGCGAGACCTGTCATTTTGCACACGGTTTGCTG 205

QY 305 gagtgcagaacactggaattcccaactgcgcacaacatagcctgagcttcattctgcag 364  
DB CATGAGCCACAACCTGATCTCACTGACCTGCCAACACAGCCTGCTTCATCTGAGCG 265

QY 365 atgtcgttcttgaagctgtgagtgaggaacagcaggggaaacgcctgtctcgaacaaca 424  
DB ACGCGGGGTAGACCTGTGGCTGGGGAACAGCAGGGGCAACACCTGGGCGCAGAGGAATC 325

QY 425 agaacacttcacatagaacgaatgattctgtggttcagttaatgatagatgagctag 484  
DB TGTACTACTCCGCCGACCTCCGTCGATTTCTGGGCTTTACGCTTTACGATGAGCTAANT 385

QY 485 ttgaacttcctgcagtgataacttaatttgcagaaaaagggcaggaagaaagatcatt 544  
DB ATGACCTTCCCGCCACATTTGACTTCATCTGAGAGAAAAAGGAGCAAGCACTACACT 445

QY 545 atgtcgttcttgaagctgtgagtgaggaacagcaggggaaacgcctgtctcgaacaaca 604  
DB ACGTTGGCCATTCACAGGCGACACACCATGTTGTCATCGCTTTTCCACCATTCACAGC 505

QY 605 tggctcagaataaataatgtatttgcctttagcaccatagccactttaaagcattgaa 664  
DB TGGCGAAGCGATCAAAACCTTCTATGATAGCTCCCGTGGCCACCGTGAAGTACACG 565

QY 665 aaagcccccgggacaaatcttgtgtcgcagatatgatgaaggaatgttctgtgca 724  
DB AAACCTGTATAACAACTCATGCTGCTCCCTGCTCTCAAGCTTATATTTGGA 625

QY 725 aaaaaaatcttctatagacacagatttctcagacaactcgttcttaacttctgtgccc 784  
DB ACAAATATTTTACCCACACACTTTTGTATCAATTTCTGCCACGACGAGGTATGCTCC 685

QY 785 agtgtattccttgatcagatttgtagtaatacatatcttctcgtgtgattcaacacca 844  
DB GGGAGACGATGATCTCTCTGACAGACGCCCTTTATCTTTGATGGATTTGACACTA 745

QY 845 acaatataaactgagccgaggaagtgatataatgcgcacacacttcctgaacacttg 904  
DB TGAACCTGACATGAGTGCTGGATGTGTATCTGTCAATATACACAGGACGATGCG 805

QY 905 tgcataatattctacactggaagcaggaatgaaatcttgcggaactccggaattgact 964  
DB TTCAGACAGTGTCTCAGCTGCTCCAGGCTGTTAACTCTGGGAGTTCACAGCTTTGACT 865

QY 965 gggggagtgagacacaaatctggaataatgcaatcagcagcaactccttgaagtcagag 1024  
DB GGGGAAGCCCACTTCAGACATGATGACTATCATCAAGCATGAGCTCCCTACTCAACCC 925

QY 1025 tcaagatatagcgttccctcagcaatgtagcagagagtgtagcagcttccaatc 1084  
DB TGACAGACATGATGATGCAATCGCATGTGGACGTTGGCAACGATTTGCTGGCCGACC 985

QY 1085 cagaagacgtgaaatgctgctcctcgtgagtgtagcacaactcactacataagaatattc 1144  
DB CTCACGATGTTGACCTTTTGGCTTTCCAGCTCCCAATCTCATTTACACAGGAAGATTG 1045

QY 1145 ctgaatgggtcagctggaattcattcattcgtggttggatgctcctcaccgatagtacaatg 1204  
DB CTCCTTACATCATCTTGACCTTATCTGCGCATGATGAGCCCTCAACCGGTTTACATG 1105

QY 1205 aaatcattcatctgatt 1221  
DB AAATGTTTCCATGATG 1122

RESULT 8  
US-09-073-674-1  
Sequence 1, Application US/09073674  
Patent No. 5998189  
GENERAL INFORMATION:  
APPLICANT: Blanchard, Claire  
APPLICANT: Benicourt, Claude  
TITLE OF INVENTION: Recombinant Dog Gastric Lipase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Warner-Lambert Company  
STREET: 2800 Plymouth Road  
CITY: Ann Arbor  
STATE: Michigan  
COUNTRY: U.S.A.  
ZIP: 48105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/073,674  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Crissey, Todd M.  
REGISTRATION NUMBER: 37,807  
REFERENCE/DOCKET NUMBER: 5072-DI-66-TMC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 734 622-1553  
TELEFAX: 734 622-1553  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1528 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double







STATE PA  
COUNTRY: USA  
ZIP: 19482  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEO for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/163,669  
FILING DATE: 30-SEP-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/852,806  
FILING DATE: 07-MAY-1997  
APPLICATION NUMBER: 60/017,915  
FILING DATE: 16-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Prestia, Paul F  
REGISTRATION NUMBER: 23,031  
REFERENCE/DOCKET NUMBER: TAR-50002-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-407-0700  
TELEFAX: 610-407-0700  
TELEX: 846189  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3271 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
OS-09-163-669-1

```

Query March 2.9% ; Score 36.6; DB 3; Length 3271;
Best Local Similarity 50.9% ; Pred. No. 0 12;
Matches 87; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Qy 969 gagtgaaccacaaaatctgtaaaaaatgcaatcagccaactctctgaaggtacagaatcag 1028
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 678 GATTGTACACAGATATATCTAGTATCAGCATTCACGAGAGCCCGCTCAGCTGTGCCAG 737

Qy 1029 agaatatcaggttcccttaacagcaaatgltgacacggaggtcagaactcttcaatccaga 1088
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 738 TGACATACAGCTTCCCAATGCGGGGCCCGGGGCGCATGAAGGACTGGGTGCGGACTCAGA 797

Qy 1089 agacgtgaaatgctgtctgtcgtcgtgaggtgacacacactcttccataatgaag 1139
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 798 GGACCGCCTCTTCTCCTGCCCAAGGAGGTCTCAGCTCTCTCCTCCACGGGAA 848

RESULT 14
US-08-276-852-154
; Sequence 154, Application US/08276852
; Patent No. 5652138

GENERAL INFORMATION:
APPLICANT: Burton, Dennis R
APPLICANT: Barbados, Carlos F
APPLICANT: Lerner, Richard A
TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES
TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS
NUMBER OF SEQUENCES: 170

CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute, Office of
ADDRESSEE: Patent Counsel
STREET: 10666 No. 5652138th Torrey Pines Road, Suite 220,
STREET: Mail Drop TPC8
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,852
FILING DATE: 18-JUL-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/178,302
FILING DATE: 30-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/954,148
FILING DATE: 30-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: SCR1452P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ ID NO: 154:
SEQUENCE CHARACTERISTICS:
LENGTH: 3282 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 15..452
US-08-276-852-154

```

```

Query Match      2.9%; Score 36.6; DB 1; Length 3282;
Best Local Similarity 51.5%; Pred. No. 0.12; Indels 0; Gaps 0;
Matches 84; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 435 catagaccagaatgagttctgggttcgtatgatagtgagttgaccttcc 494
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 302 CACAGCCTCAGTGAAGTGTAGAGAGCTCAGGCTGCAGACAGCGCTTATTATGTGC 361

QY 495 tgcagtgataaacttatttgcagaanaacgggcaggaagaatctatattgcgcta 554
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 362 GAGAGGGGGGCCATATAGTTGGGATGATCTCCCGACAGCAATTATTATGACGCTCG 421

QY 555 ttcaaggccaccacatgggttttgcatttcaccatg 597
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 422 GGGCAAGAGAACCGCGTCATCTGAGCTCAGCTCCACCAAG 464

```

```

RESULT 15
US-08-276-852-169/c
; Sequence 169, Application US/08276852
; Patent No. 5652138
; GENERAL INFORMATION:
; APPLICANT: Burton, Dennis R
; APPLICANT: Barbas, Carlos F
; APPLICANT: Lerner, Richard A
; TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES
; NUMBER OF SEQUENCES: 170
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; ADDRESS: Patent Counsel
; STREET: 10666 No. 5652138th Torrey Pines Road, Suite 220,
; CITY: Mail Drop TPC8
; STATE: La Jolla
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,852
FILING DATE: 18-JUL-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/178,302
FILING DATE: 30-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/954,148
FILING DATE: 30-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: SCR1452P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ ID NO: 169:
SEQUENCE CHARACTERISTICS:
LENGTH: 3282 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-276-852-169

```

```

Query Match      2.9%; Score 36.6; DB 1; Length 3282;
Best Local Similarity 51.5%; Pred. No. 0.12;
Matches 84; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 435 catagaccagaatgagttctgggttcgtatgatagtgagttgaccttcc 494
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2981 CACAGCCTCAGTGAAGTGTAGAGAGCTCAGGCTGCAGACAGCGCTTATTATGTGC 2922

QY 495 tgcagtgataaacttatttgcagaanaacgggcaggaagaatctatattgcgcta 554
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2921 GAGAGGGGGGCCATATAGTTGGGATGATCTCCCGACAGCAATTATTATGACGCTCG 2862

QY 555 ttcaaggccaccacatgggttttgcatttcaccatg 597
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2861 GGGCAAGAGAACCGCGTCATCTGAGCTCAGCTCCACCAAG 2819

```

Search completed: January 30, 2002, 22:11:14  
Job time: 6251 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2002, 20:23:58 ; Search time 1536.27 Seconds  
(without alignments)  
8876.300 Million cell updates/sec

Title: US-09-578-063-46

Perfect score: 1269  
Sequence: 1 acgttggaacacctgtcaag.....gacgctgtgagccgcatatg 1269

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthm:\*  
3: em\_estlm:\*  
4: em\_estlm:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hlc:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_hlc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rnd:\*  
20: em\_gss\_vit:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	618.2	48.7	2927	12 AK019504	AK019504 Mus muscu
2	505	39.8	954	11 BF135102	BF135102 601780142
3	497.6	39.2	1079	11 BF141567	BF141567 601788654
4	471.8	37.2	669	11 BF143546	BF143546 601790896
5	461.8	36.4	1350	12 AK010093	AK010093 Mus muscu
6	460.2	36.3	1347	12 AK010116	AK010116 Mus muscu
7	460.2	36.3	1349	12 AK009300	AK009300 Mus muscu
8	460.2	36.3	1350	12 AK010139	AK010139 Mus muscu
9	458.6	36.1	1340	12 AK010124	AK010124 Mus muscu
10	458.6	36.1	1343	12 AK009479	AK009479 Mus muscu
11	458.6	36.1	1345	12 AK009413	AK009413 Mus muscu
12	458.6	36.1	1345	12 AK009473	AK009473 Mus muscu

13	458.6	36.1	1346	12 AK009359	AK009359 Mus muscu
14	458.6	36.1	1346	12 AK009523	AK009523 Mus muscu
15	458.6	36.1	1346	12 AK009729	AK009729 Mus muscu
16	458.6	36.1	1348	12 AK010058	AK010058 Mus muscu
17	458.6	36.1	1349	12 AK009474	AK009474 Mus muscu
18	458.6	36.1	1349	12 AK009546	AK009546 Mus muscu
19	458.6	36.1	1349	12 AK009571	AK009571 Mus muscu
20	458.6	36.1	1349	12 AK009573	AK009573 Mus muscu
21	458.6	36.1	1349	12 AK010148	AK010148 Mus muscu
22	458.6	36.1	1350	12 AK009459	AK009459 Mus muscu
23	458.6	36.1	1350	12 AK009525	AK009525 Mus muscu
24	458.6	36.1	1350	12 AK009773	AK009773 Mus muscu
25	458.6	36.1	1350	12 AK010061	AK010061 Mus muscu
26	458.6	36.1	1350	12 AK010125	AK010125 Mus muscu
27	458.6	36.1	1350	12 AK010203	AK010203 Mus muscu
28	458.6	36.1	1351	12 AK009428	AK009428 Mus muscu
29	458.6	36.1	1352	12 AK009431	AK009431 Mus muscu
30	458.6	36.1	1356	12 AK010019	AK010019 Mus muscu
31	458.6	36.1	1356	12 AK010035	AK010035 Mus muscu
32	457.6	36.1	1346	12 AK009437	AK009437 Mus muscu
33	457.6	36.1	1347	12 AK010103	AK010103 Mus muscu
34	457.6	36.1	1349	12 AK010026	AK010026 Mus muscu
35	457	36.0	1346	12 AK009560	AK009560 Mus muscu
36	457	36.0	1347	12 AK010236	AK010236 Mus muscu
37	457	36.0	1348	12 AK010231	AK010231 Mus muscu
38	457	36.0	1349	12 AK009875	AK009875 Mus muscu
39	457	36.0	1350	12 AK009537	AK009537 Mus muscu
40	456	35.9	1349	12 AK010106	AK010106 Mus muscu
41	455.4	35.9	1345	12 AK009544	AK009544 Mus muscu
42	455.4	35.9	1349	12 AK009990	AK009990 Mus muscu
43	453.8	35.8	1346	12 AK010173	AK010173 Mus muscu
44	449.8	35.4	1407	12 AK009923	AK009923 Mus muscu
45	448.2	35.3	1350	12 AK010123	AK010123 Mus muscu

#### ALIGNMENTS

RESULT 1  
LOCUS AK019504  
DEFINITION Mus musculus 0 day neonate skin cDNA, RIKEN full-length enriched library, clone:4632427C23, full insert sequence.  
ACCESSION AK019504.1 GI:12859754  
VERSION AK019504.1  
KEYWORDS CAP trapper.  
SOURCE Mus musculus (strain:C57BL/6J) 0 day neonate skin cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library  
clone:4632427C23.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 2927)  
REFERENCE Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Methods in enzymology. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
REFERENCE 2 (bases 1 to 2927)  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, N., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome research. 10 (10), 1617-1630 (2000)  
MEDLINE 10499374  
PUBMED 11042159  
REFERENCE 3 (bases 1 to 2927)  
AUTHORS Shibata, K., Itoh, N., Aizawa, K., Nagaoke, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, D.,

```

TITLE
Oikazaki,Y., Murmatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
Riken integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome research. 10 (11), 1757-1771 (2000)

JOURNAL
20530913
MEDLINE
11076861
PUBMED
4 (bases 1 to 2927)
REFERENCE
The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5 (bases 1 to 2927)
AUTHORS
Aachai,T., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
Atakawa,T., Carrinchi,P., Fukuda,S., Fukunishi,Y., Funuo,M.,
Haneagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Horl,F.,
Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,T.,
Kojima,Y., Konno,H., Kouda,M., Koya,S., Kutharra,C., Matsuyama,T.,
Miyazaki,A., Nishii,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
Sasak,D., Shibata,K., Shibata,Y., Shitagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,Y.,
Yoshida,K., Yoshino,M., Murmatsu,M. and Hayashizaki,Y.

TITLE
Direct Submission
JOURNAL
Submitted (18-AUG-2000)
COMMENT
Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
CNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGAGAGAGATTCCTGACGTAAATTAATTAATCCCCCCCCCCC 3']. cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. cDNA wenthru
through one round of normalization to Rot = 10.0 and subtraction to
Rot = 100.0. Second strand cDNA was prepared with the primer
GAGAGAGAGATTCCTGACGTAAATTAATTAATCCCCCCCCCCC 3']. cDNA was cleaved
with BamHI and XhoI. cDNA of size comprised between 3 and 7 kb was
selected before cloning. Vector: a modified p Bluescript KS(+) after
bulk excision from lambda FLC I. Cloning sites, 5' end: Salt; 3'
end: BamHI. Host: DH10B.
Location/Qualifiers
1..2927
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGI:1912763"
/db_xref="MGI:1926003"
/cross="4632427C23"
/tissue="skin"
/clone_id="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
153..662
/note="putative"
/codon_start=1
/protein_id="BA81766.1"
/db_xref="GI:12859755"
/translation="MSEITLSRWTVSRREIWLILVAYLQNRVNSGHLPTKADPEE
AFNWSNIIIRKAGPSESEVATEDGVIILSVNRIPROGTRIKKSGSRPVYLQGLLG
DASNWSIINIKSGFIADAGFDVWNGNSRGNTWSRKHRTLSDIDDEFNAFRKNTLL
RVEDEIDMP"
2908..2913
/note="putative"
2927
/note="putative"

```

Query Match	48.7%	Score 618.2	DB 12	Length 2927
Best Local Similarity	85.3%	Pred. No. 2.1e-169		
Matches 702	Conservative 0	Mismatches 118	Indels 3	Gaps 1
447	tgagtcctgagcttcacgataatgatagagatgctggtttgaccttcctccgacgagatataa	506		
Db	1376	ttctctttcccttcttctgattatgatgaataatgactgaggtttcccttccacgtctgattata	1435	
Qy	507	cttatacttcgcaaaacgagcgccaggaanaagaatctataatgctggtctatccacagggcac	566	
Db	1436	ctttatctttacagaaacggccgacagaaaaaggtctattatgtgcgctactccacagggcac	1495	
Qy	567	caacatgggtcttattatgcatcttccacatgcgcagagctggtcgcagaaataaataatgta	626	
Db	1496	caccatggcgcttattatgcatcttccacatgcgcagagctggtcgcagaaataaataatgta	1555	
Qy	627	tttctcttaagacccatagccactgcttaagcatgcaaaagccccggacccaattttt	686	
Db	1556	tttttgccttacctcttattatgacccactgcttataatgcttccacagaaatcttct	1615	
Qy	687	gttctgctcagaaatgatgatcaagggatgtgttggcaaaaaagaatttcgtatcagac	746	
Db	1616	gctgctgcacagatgatgatcaagggatgtgttggcaaaaaagaatttttttttaccagac	1675	
Qy	747	cagattctcagacaaactgttattactcctgttgccgagcggtatcttgatcagattg	806	
Db	1676	ttgattttttccacacagcttt	1735	
Qy	807	tagtaatacatgttactctctgtgtgtatcaacacccaacaaatgatgaatgacgagc	866	
Db	1736	cagcaacatcatcttattcttctgtgtgtatcaacacccaacaaatgatgaatgacgagc	1795	
Qy	867	aagtgatatactgtcccaacactctgtcggaaacatctgtgcaaaatatcttctactggg	926	
Db	1796	aaatgtgatgtgtgcgccacacacacgctgtaagaaacgctgtgcagaaacattctccactggag	1855	
Qy	927	ccagcgaatgaaatctctgtgaaactccgggcatcttactgtgggagtgagacccaanaatc	986	
Db	1856	ccagcgcctggaattctgtgggaactctgtgccccttgcactgggaatgagaaacccaanaatc	1915	
Qy	987	ggaataatgcaatcagcccaactcctgttaaggtacagagatcagagatagaaggtccctac	1046	
Db	1916	ggagaaatgtatgtacgccaacacctcttataaggtatccagatattatgtgtcccaac	1975	
Qy	1047	agcaatgtggaaggaaggttcagactgcttcaaatccagaagacgtgaaatgctgct	1106	
Db	1976	agcaatgtggaaggaaggttcagactgcttcaaatccagaagacgtgaaatgctgct	2035	
Qy	1107	ctctgaagtgcaaacactcatcattacataaataatctcgtgaatgggtcagctgattt	1166	
Db	2036	ttctgaagtgcaaacactcatcattacataaataatctcgtgaatgggtcagctgattt	2095	
Qy	1167	catctggggttggatgtctctccacacggtatgaacatgaatcatcatctatgacga	1226	
Db	2096	catctggggttggatgtctctccacacggtatgaacatgaatcatcatctatgacga	2154	
Qy	1227	ggaagagaccaaacttcccaaggaacggtgtgtagacgcgtatg	1269	
Db	2155	--AGGAGGCCAAACCTTCCCGCAGGAACCTGCAGGGTCAAAATTG	2195	

RESULT 2

BF135102

954 bp

EST

24-OCT-2000

LOCUS

601780142P1 NCI\_CGAB\_Lu30 Mus musculus cDNA clone IMAGE:4008131 5'

DEFINITION

mRNA sequence.

ACCESSION

BF135102.1

GI:10974142

VERSION

EST

KEYWORDS

SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 954)  
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Plate: LLAM9242 row: c column: 12  
High quality sequence stop: 695.

FEATURES  
Location/Qualifiers  
1..954  
/organism="Mus musculus"  
/strain="CZECH II"  
/db\_xref="taxon:10090"  
/clone\_1db="NCI-CGAP\_Lu30"  
/clone\_1lb="NCI-CGAP\_Lu30"  
/tissue\_type="tumor, metastatic to mammary"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pCMV-SPORT6; Site: 1; NotI; Site: 2; SalI; transgenic model MMTV-1, expression driven by MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 280 a 222 c 222 g 230 t  
ORIGIN

Query Match 39.8%; Score 505; DB 11; Length 954;  
Best Local Similarity 85.3%; Pred. No. 1.9e-136;  
Matches 598; Conservative 0; Mismatches 100; Indels 3; Gaps 3;

QY 493 cctgcagtgataaacttatttgcgagaaaagcgagcaagaatctattatgcgcg 552  
DB 1 CCAGCTGTGATAAAGCTGATCTTACAGAAAAGCGGCGAAGAAAGCTATTATGTCGCGC 60

QY 553 tattcacaggcgaccacatgggcttattgcatttccaccatgcagagctggctcag 612  
DB 61 TACTCACAGGCGACACCATGGGCTTATGCAATTTCCCAATGCCAGAGCTAGCTCAT 120

QY 613 aaaaacaaatgtatttcttagcacccatagcactgttaagcatgcaaaaagcccc 672  
DB 121 AAATCAAAATGTATTGTGCTTAGCTCCATATACCACTGTATAATGCAAGAAGTCTT 180

QY 673 gggaacaaatcttctgtgcgcagagataatgataagagatgtttgcgaagaagaa 732  
DB 181 GGTAACCAAAATTTCTGCTGTCGACAGATGATGATGATTTTGGCAGACAAAG 240

QY 733 ttctgtatagacagcaatttccagacaactgttattaccttctgtgcagagttat 792  
DB 241 TTTTATACAGACTAGATTTTTCAGGACAGCTTTTATTTACCTTGTGCGCAATGATT 299

QY 793 ctgtatcagatttgtagtaataatcatgtactcttctgtgtgattcaacaacaatc 852  
DB 300 CTCGACCAAAATCTGCACCAACATCTTACTCTGGAGAGTTTAAACAAACAAATATG 359

QY 853 aaatagcgagcgaggaagtgtatatgtcgcgcacactcttctgtgaacaatctgtc 912  
DB 360 AACATGACAGAGCAAAATGTGTATGTCGCCATACACCTGCTGGAACGCTGTGTCAGAAC 419

QY 913 attctaacctggagcgaggaagtgaattctgtgtgaatccgggtgattgagctgtg 972  
DB 420 ATTCTCCACACGAGGAGCCAGGCGGTGAATCTGCTGCTTTGATGGGGAAGT 479

QY 973 gagacaaaaatctggaaaaatgcatcagccactcctgttaaggtacagatcagat 1032  
DB 480 GAGACCAAAATTCAGGAGAAATGTAAATCAGCCACTCTTAAAGTTCAAGTTCCGAGAT 539

QY 1033 atgacggtcccttaagaatctgagcagagagtgtagcagcttctcaatccagaagac 1092  
DB 540 ATGATGGTCCCAACAGCAATGTGACGTGAGGCTCAAGACTGG-TTCAATCCAGATGAT 598

QY 1093 gtgaanaatgctgtctctctgaggtgacaaacctcatctacaataaattctgtatg 1152  
DB 599 GTGAACAAATTAATCTTCTGAAGTAAACCAACTCATCTACACAGAACATCTCGAGCTG 658

QY 1153 gctcagtgatctcactcctgaggttgtagtgcctaccg 1193  
DB 659 GGTCAATGTGA-TTCATCTGGGGGTGATGCCCTCAGCG 698

RESULT 3  
LOCUS BF141567 1079 bp mRNA EST 24-OCT-2000  
DEFINITION 601788654F1 NCI-CGAP\_Lu30 Mus musculus cDNA clone IMAGE:4016413 5', mRNA sequence.  
ACCESSION BF141567  
VERSION BF141567.1 GI:10980607  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 1079)  
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Plate: LLAM9263 row: 1 column: 14  
High quality sequence stop: 663.

FEATURES  
Location/Qualifiers  
1..1079  
/organism="Mus musculus"  
/strain="CZECH II"  
/db\_xref="taxon:10090"  
/clone\_1db="NCI-CGAP\_Lu30"  
/clone\_1lb="NCI-CGAP\_Lu30"  
/tissue\_type="tumor, metastatic to mammary"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pCMV-SPORT6; Site: 1; NotI; Site: 2; SalI; transgenic model MMTV-1, expression driven by MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 312 a 267 c 263 g 237 t  
ORIGIN

Query Match 39.2%; Score 497.6; DB 11; Length 1079;  
Best Local Similarity 84.1%; Pred. No. 2.8e-134;  
Matches 621; Conservative 0; Mismatches 109; Indels 8; Gaps 5;

QY 256 aagaagcaggttcagagcgtgtgttctactcagcatgtgctagtgtgaggtgcagc 315  
DB 20 AAGAAGAGAGAGATCCAGGCGAGTGTCTACTGCAACATGCTTCTTGGGAGATGAC 79

QY 316 aacgtgattccaaacctgcgcacaacaatagctgtgctcattctgtgagatgtctgt 375  
DB 80 AACTGATTTCCACACCTGCGCCACACAGCGCTGGGTTTATTCTGCGAGATGACGTTT 139







```

Db 672 AAATCCCTTTAAAGATTTTCACTTTCCTAGTTCCTTCACAGTGTATTTTGTA 731
Oy 725 aaaaagaattctgtatcagaccagattctcagacaactgttattacccttggcc 784
Db 732 ACAAAATGTCATGCCCCACAACTACTATCATATTTCTTGGTACGGAAGTGTCTCAC 791
Oy 785 aggtgattctgtatcagattgtatgatatcatctgttaccctctgggtggattcaacc 844
Db 792 GGGGAGCTGTAGATCTTCTGTGACGACACCTTTATTCATCTTGTGTGATTTGACAGA 851
Oy 845 acaatgatcaatgagccgagcaagtgtatcatctgtcccaactctgtcgtgagacactg 904
Db 852 AAAAAGTAAATGTGAGTGGCTGTGATGTATCATAGGCAATTAACGACGAGAACATCTA 911
Oy 905 tgcgaatattctacactgtgagccagcagtgaaattctgttgaactccgggacattgact 964
Db 912 CTCAGAGACCTTTTCCACTGGGACACGCTTAAATCTGGGAAGCTTCAAGCCCTATTA 971
Oy 965 gggggagtgagcccaaaatctgaaatgcaatcagcccaactccctgtaagtgacag 1024
Db 972 GGGGAGCTGTAGATCTTCTGTGACGACACCTTTATTCATCTTGTGTGATTTGACAGA 1031
Oy 1025 tcaagagatgagcgtccctacagcaatgtgagacagagtgcaagactgttcaaatc 1084
Db 1032 TGTGACGCAATGACCGTGTGCAATTCAGTGTGGAACGCTGCGCATGATCTGCTGATC 1091
Oy 1085 cagaagacgtgaaatgctgtctctgtgaggtgagcaaccatctacataagaatattc 1144
Db 1092 CCCAAGATGTGCAATGCTGCTGCTGCAATGCTGCAATGCTGCAATGCTGCAATGCTG 1151
Oy 1145 ctgaatgtgagcagcgtgattcattcctgtgggtgttggatgtctcaccgtatgtacatg 1204
Db 1152 TTCCCTCAATGACCTGAGTCTCATCTGCGGATGATGCGCTCAAGAGGATTTACATG 1211
Oy 1205 aatcatcatctgattg 1221
Db 1212 AGATAGTTACCATGATG 1228

```

**RESULT 6**  
**LOCUS** AK010116 1347 bp mRNA HMC 05-JUL-2001  
**DEFINITION** Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310068005, full insert sequence.  
**ACCESSION** AK010116  
**VERSION** AK010116.1 GI:12845334  
**KEYWORDS** CAP trapper.  
**SOURCE** Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone:11b:RIKEN full-length enriched mouse cDNA library  
**ORGANISM** Mus musculus  
**REFERENCE** 1 (bases 1 to 1347)  
**TITLE** Carninci, P. and Hayashizaki, Y.  
**AUTHORS** Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
**JOURNAL** Methods in enzymology. 303, 19-44 (1999)  
**MEDLINE** 99279253  
**PUBMED** 10349636  
**REFERENCE** 2 (bases 1 to 1347)  
**TITLE** Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
**AUTHORS** Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
**JOURNAL** Genome research. 10 (10), 1617-1630 (2000)  
**MEDLINE** 20499374  
**PUBMED** 11042159  
**REFERENCE** 3 (bases 1 to 1347)  
**TITLE** Shibata, K., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

```

Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, F., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
JOURNAL RIKEN integrated sequence analysis (RISA) system-384-format
REFERENCE sequencing pipeline with 384 multichipillary sequencer
GENOME Genome research. 10 (11), 1757-1771 (2000)
GENOME 20530913
PUBMED 11076861
REFERENCE 4 (bases 1 to 1347)
TITLE The RIKEN genome exploration Research Group Phase II Team and the
AUTHORS FANTOM Consortium.
JOURNAL Functional annotation of a full-length mouse cDNA collection
REFERENCE Nature 409, 685-690 (2001)
AUTHORS 5 (bases 1 to 1347)
TITLE Aizawa, K., Akahira, S., Akiyama, T., Aono, H., Arai, A.,
AUTHORS Arakawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M.,
JOURNAL Hanagaki, T., Hara, A., Hayatsu, N., Himoto, K., Hirooka, T., Horii, F.,
REFERENCE Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J.,
AUTHORS Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T.,
JOURNAL Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y.,
REFERENCE Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H.,
AUTHORS Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T.,
JOURNAL Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
REFERENCE Tanaka, T., Teijima, Y., Toya, T., Yamamura, T., Yasunishi, A.,
AUTHORS Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
REFERENCE Physical and Chemical Research (RIKEN), Laboratory for Genome
AUTHORS Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
JOURNAL RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
REFERENCE Kanagawa 230-0045, Japan (E-mail:genome-res@gsr.riken.go.jp,
AUTHORS URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
JOURNAL Fax: 81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for
  further details.
FEATURES
  source
    Location/Qualifiers
      1..1347
        organism="Mus musculus"
        strain="C57BL/6J"
        db_xref="taxon:10090"
        /db_xref="MGI:1895801"
        /db_xref="MGI:1914967"
        /clone="2310068005"
        /sex="male"
        /tissue="tongue"
        /clone_11b="RIKEN full-length enriched mouse cDNA library"
        /dev_stage="adult"
        51..1238
        /note="putative"
        /protein_id="BAB26711.1"
        /db_xref="GI:12845335"
        /translation="MTLLVTVLSAFCGAGHGFEGKLGKPNPEANNNVSOMITVGYC
        SERIEVYEDGIIILGYRITPTGKNSSENGKRPVATLQGLASATNWTNPNNSIA
        FLIADGVDVWLGNSRGNWTSRKNVYSPDSVFMAFSEDEAKDLDLPATDIPVOVK
        GCEKIHVGHSGGTIGFATSTNPALAKIKRFYLAIVAVYKTESPCKIISLPK
        FLKIVFGKMPHNYLDIPLGTETVSCSELDLDCSNLFTFCGDKNLNVSRGV
        YLGHNPAGSTDDLFHMAOLAKSGKQIAVNMWSPLOMNAHYNOKPPYDVASMTYPI
        AVWNGCHDILADPQVAMLLPLKPNLYHKELIPYHDLFIAMDAPOEVNIEITYTM

```







Db	792	GGGACCTCTTAATATTTCTCTGACACAAAGCTTTATTCATCTTCTGTGGATTGCACAGA	851
Qy	845	acaatataaacatgagccgagcaatgataatgctgcccaactcttgctggaactctg	904
Db	852	AAAACCTTAAGTGTAGTGCCTTTGATGTGTATCTAGGGGATATATCCAGCAGAACATCTA	911
Qy	905	tgcacaaatattctaacatctggaacccaggccagtgtaattcttgctgtaacctggacttgc	964
Db	912	CTCAAGACCTTTTCCACTGCGGACAGCGCTTCGTAAATCTGTGGAAAGCTTCACAGCCTTAACT	971
Qy	965	gggggagtgatgagacccaataatctggaataaatgcaatcagccaactctgtaagttacagag	1024
Db	972	GGGGAGAGTCCATTACAGAACATGTATACATACAAATTCAGAAAAGCGCTCCCTACTATGATG	1031
Qy	1025	tcaagatatatacaggtctccctacagaatgtaatgtgacagggaggtgaagactggtcttcaatc	1084
Db	1032	TGTCAAGCATATGACCTGCGCAATTGCAAGTGTGGAAACGATGGCCATGACATCTTGCTGATC	1091
Qy	1085	cagaagacgctgaaatgtctgtctctctgaggtgacccaactctacataagaataatc	1144
Db	1092	CCCAAGATGTGCGCAATGCTGCTTCCCAAACTCCCAACCTTTGTACATTAAGAGATTC	1151
Qy	1145	ctgaatgggctcagctgagttcatctacatctgggtgttgatgtctctcctacgtaagtacaatg	1204
Db	1152	TTCCCTACATATCTCTTGAGCTTTCACTGTGGCGCATGGATGATCGGCTCAAGAGTTTACAATG	1211
Qy	1205	aaatcatccatctgagt 1221	
Db	1212	AGATGTATACCATGATG 1228	
RESULT 9			
LOCUS	AK010124	AK010124 1340 bp mRNA HTC 05-JUL-2001	
DEFINITION		Mus musculus adult male tongue cDNA, RIKEN full-length enriched	
ACCESSION		library, clone:2310069D09, full insert sequence.	
VERSION		AK010124	
KEYWORDS		AK010124.1 GI:12845346	
SOURCE		CAP trapper.	
ORGANISM		Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone:11b:RIKEN full-length enriched mouse cDNA library clone:2310069D09.	
REFERENCE		Mus musculus	
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
TITLE		1 (bases 1 to 1340)	
JOURNAL		Carninci, P. and Hayashizaki, Y.	
MEDLINE		High-efficiency full-length cDNA cloning	
PUBMED		Methods in enzymology. 303, 19-44 (1999)	
REFERENCE		99279253	
AUTHORS		10349636	
TITLE		2 (bases 1 to 1340)	
JOURNAL		Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,	
MEDLINE		Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	
PUBMED		Normalization and subtraction of cap-trapper-selected cDNAs to	
AUTHORS		prepare full-length cDNA libraries for rapid discovery of new genes	
TITLE		Genome research. 10 (10), 1617-1630 (2000)	
JOURNAL		3 (bases 1 to 1340)	
MEDLINE		Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,	
PUBMED		Kono, H., Akiyama, T., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,	
AUTHORS		Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,	
TITLE		Yamanoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,	
JOURNAL		Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,	
MEDLINE		Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J.,	
PUBMED		Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.	
AUTHORS		RIKEN integrated sequence analysis (RISA) system--384-format	
TITLE		sequencing pipeline with 384 multiplexillary system	
JOURNAL		Genome research. 10 (11), 1757-1771 (2000)	
MEDLINE		20530913	
PUBMED		11076661	

[illegible]











DB	911	CTCAAGACCTTTTCCACGTGGGACACGCTTGCTTAATCTGGGAGACTTCAAGCCTTAACCT	970
OY	965	ggggaggtgagagcccaaaaactctggaataatgcaatcagcccaactcctgtaaggtacagag	1024
DB	971	GGGGAAGTCCATTACAGAAACATGTTACACTCAATCAATCAAGAAACGCCCTCCCTACTATGATG	1030
OY	1025	tcagagatatgacggtccctcagcaaatgtgtgacagaggtctcaggaactcgttcttaaac	1084
DB	1031	TGTCAGCCATGACCTGGCCAAATGTGACGTGTGACGCGCCATCCTCGCTGATG	1090
OY	1085	cagaagacgtgaataatgctcgtcctctggtggtgacacactcattaccataagaatattc	1144
DB	1091	CCCAAGATGTGGCAATGCTGCTTCCCAACTCCCAACCTTTGTATACCATAGGAGATTC	1150
OY	1145	ctgaatggcctcagctggtattcattcattcgtgggtgtgattgcatcaccgtatgtacaat	1204
DB	1151	TTCCTACAAATCACCCTGGACTCATCTGGCGCATGGATGCGGCTCAAGAGTTTCAATG	1210
OY	1205	aaatcattccattcattg	1221
DB	1211	AGATGATTACCATGATG	1227
RESULT	12		
AK009473			
LOCUS	AK009473	1345 bp	mRNA
DEFINITION			HTC
ACCESSION			05-JUL-2001
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			

FEATURES	Source
1. 1345	Location/Qualifiers
/organism="Mus musculus"	
/strain="C57BL/6J"	
/db_xref="taxon:10090"	
/db_xref="MGD:MG1:1898391"	
/db_xref="MGD:MG1:1914967"	
/clone="2310022H17"	
/sex="male"	
/tissue_type="tongue"	
/clone_1ib="RIKEN full-length enriched mouse cDNA library"	
/dev_stage="adult"	
52. 1239	
/note="putative"	
/codon_start=1	
/protein_id="BAB26312.1"	
/db_xref="GI:12844294"	
/translation="MWLLIVTSVLSAFGAGHLRGKLGPKKPEANMNVSMQITWGVPR SEEFVYTEDCYILGVYRIRIPGKNSENIGRPAVYQHGILASATWITLPLNLSLA FTLADAGDWLGNRSKGTWSRKNNVYSPDSVEAWFSDENATDLPATIDFVQRT GQELIHVHGSHGSGCTTIGFIASTFPAALAKIKRPAALAPATVATVTSPEFKSLIPK FLNGVHGNKMFPMHNYLDOLFGEVCSRELLDGLGNALTEFCGPKKNLANSRPFVD YILGNPAGTSTODPEHMAOLAKSGTLOAYWMSPLQNNLHYNOKTPEYDVSAATVDFD AYNMGCHDILADPDQVAMLPKPLENLHYKLEILFYNLDFIMADADQEVYNELVTMW AED"	
BASE COUNT	395 a 288 c 283 g 379 t
ORIGIN	
Query Match	36.1% Score 458.6; DB 12; Length 1345;
Best Local Similarity	63.6% Pred. No. 7.3e-123;
Matches 698; Conservative	0; Mismatches 399; Indels 0; Gaps 0;
Oy	125 acccgaagcattcatgataataatgaatcatccaacatcaaggctacccctgtgagc 184
Db	131 ACCCGAAGCAACACTGATGTATGTCAGATGATTAACCTACTACGGGAGATATCCAAAGTGAGG 190
Oy	185 aatatgaatgcgaactgaagatggatataccctcttcgtttaacaggatccctcgaggcc 244
Db	191 AATATGAAGTGTATCTGCTAAGATGGCTCAATCTTGCGGGGTCTATGAAATTCCTTATGGGA 250

QY 245 tagtgaacctagaagaagaggttccagagcctgtgttactgcagctgtgcctgtg 304  
 Db 251 AGAAAAATTTCTGATATATGCGCAAGAGACGTGTGCGCATATTTGCGCATATGTTGATG 310  
 QY 305 gagggtgctgacagaggtttccaaactgcccacaatgctggcctgctatcttgcgag 364  
 Db 311 CATTACGCCAANAACGTGGATTACAAATCTGCCAANAACAGCGCTGCGCTTATTCTAGCAG 370  
 QY 365 atgtgtgttttgcagtggtgagtgaggaagacagaggggaaagcgtgtctccgaacaaca 424  
 Db 371 ATGCTGTGCGATATGATGTGTGCGTGGGGAACAGTGCAGGAATATCATGTCTCCGAAAAATG 430  
 QY 425 agacactctccatagaaccaagatgagttctgtggcttccagttatgagtgtgcctagt 484  
 Db 431 TATATATTCCACAGACTGATGTAATTTCTGGGCTTTACAGCTTTGATGAATGGCTAAAT 490  
 QY 485 ttgacctctctcagtgatgataaaacttatttttcagaaacggcgccggaagaagatctat 544  
 Db 491 ATGACCTTTCAGCCACCATATGACTTCTATTGTACGAAAACTGGACAGAAATATACACT 550  
 QY 545 atgtcgtctatccacagagggcacaccatggtcttattgcatttccacatgcccagagc 604  
 Db 551 ATGTTGTGCTACTCTCAGGGACACCATATGGGTTTATGTGCTTTTCTTACCAATCTGCTC 610  
 QY 605 tggctcagaagaatacaaatgtatatttgccttagcacccatagccactgttaagatgcaa 664  
 Db 611 TGGCTTAAAAAATCAAGAGGTTTATGATGATGCTTACAGTTGCTTACTGTAAGTATACAG 670  
 QY 665 aaagcccgaggacaaatttggctgtcgcagatagatgataaggaattgtttggca 724  
 Db 671 AATGCTCCCTTTAAAAAGATTTCATATTCCTTAAGTTTCTTCCAAAGGTGATATTTGGTA 730  
 QY 725 aaaaagaattctgtatagacagacagattctccagacactgttatttacccttgtgagcc 784  
 Db 731 ACAAAATGTTCAATGCCCAACACTACTTATGATCAATTTCTGTGAGGAAGTGTGCTCAC 790  
 QY 785 aggtgattcttgatcagatttgatgataataatgatttacttctgtgtgagtaaccaca 844  
 Db 791 GGGAGTGTGATATCTTCTCTGACGACAGCGTTTATTCATCTTGTGATTTGACAAAGA 850  
 QY 845 acaatatgacatgagccgaagaagtgatagctgtccacactctgtgtgagaactcg 904  
 Db 851 AAAAATTAAATGTGAGTCTGTTGATGTATCTAGGCGATTAATCCAGCAAGCAATCTA 910  
 QY 905 tgcgaataatcttaacatgagcagcagcagtgtaattctgtgtaactccgggacttgact 964  
 Db 911 CTCAAGACCTTTTCCACTGTGGCAGAGCTGTAAATCTGGGAAGCTTCAAGCCTATAACT 970  
 QY 965 gggggagtgtagacaaataatctggaabaatgcaatcagccaactcctgtaaggtagagag 1024  
 Db 971 GGGGAGTGTCAATTACAGCAATGTTTACACTTAATCAGAAAGCGCTCCCTACTATGANG 1030  
 QY 1025 taagagatgtagcgtgtccctacagcaatgtgacagaggaagtgaagctgtgttcaatc 1084  
 Db 1031 TGTCAAGCCATGACCGTGGCAATTTGCAAGTGTGAACGCTGGCAATGCAATCTGCGTATC 1090  
 QY 1085 cagaagaagctggaataatgtcgtctctctgaggtgacacacctatccataagaatatctc 1144  
 Db 1091 CCCAAGATGTGCCAATGCGCTGCCCAAACTCCCAACTTCTGTACATTAAGAGATATC 1150  
 QY 1145 ctgaatgggtcagtgatgattcattcattctgtgggtttgagatgcttcaacgtagataatg 1204  
 Db 1151 TTCCTTACATACCTGTGACTTCACTGTGGCGATGATGCGCTCAAGAGTTTACAATG 1210  
 QY 1205 aatcatcatcatgagt 1221  
 Db 1211 AGATAGTTACCATGATG 1227

RESULT 13  
 AK009359  
 LOCUS AK009359 1346 bp mRNA HTC 05-JUL-2001

DEFINITION  
 Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310015H15, full insert sequence.  
 ACCESSION  
 AK009359  
 VERSION  
 AK009359.1 GI:12844108  
 KEYWORDS  
 CAP trapper.  
 SOURCE  
 Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone:11b:RIKEN full-length enriched mouse cDNA library  
 clone:2310015H15.  
 ORGANISM  
 Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 Carninci, P. and Hayashizaki, Y.  
 TITLE  
 High-efficiency full-length cDNA cloning  
 JOURNAL  
 Methods in enzymology. 303, 19-44 (1999)  
 PUBMED  
 99279253  
 10349636  
 2 (bases 1 to 1346)  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 TITLE  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL  
 Genome research. 10 (10), 1617-1630 (2000)  
 2049374  
 11042159  
 3 (bases 1 to 1346)  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 TITLE  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichipillary sequencer  
 JOURNAL  
 Genome research. 10 (11), 1757-1771 (2000)  
 20350913  
 11076861  
 4 (bases 1 to 1346)  
 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
 TITLE  
 Functional annotation of a full-length mouse cDNA collection  
 JOURNAL  
 Nature 409, 685-690 (2001)  
 5 (bases 1 to 1346)  
 Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Hanagaki, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-research.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
 COMMENT  
 Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTCTTTTCA 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went

through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGAGATTCTCGAGTAATTAAATTAATCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.  
Host: SOLR.

## FEATURES

## Source

Location/Qualifiers  
1. 1346  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/db\_xref="MGI:1895805"  
/db\_xref="MGI:1914967"  
/clone="2310015H15"  
/sex="male"  
/tissue\_type="tongue"  
/clone\_id="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
53. 1240  
/note="putative"  
/codon\_start=1  
/protein\_id="BAB26240.1"  
/db\_xref="gi:12844109"  
/translation="MWLLIVSVLSAFGAGLGGKLGPKPEANMANSOMITWGP  
SEEVETEDYILGVRIPIYGRKNSENIIGRPVAYLQHLASKNITMLPNNSLA  
FILDAGYDWMGNSRNTWSRKNVYSPDSVEAFSEFDMAXKDLPAIDIFVOKA  
GOEKIHYVSGGTIGFIAFTSNPALMKIRFALAPVATVKTSEPFKIHIFPK  
FLKVIENGKEMPHNYLDFLGTGVECSRELIDLSALFIPGSEDKNNSRFDY  
YLGNPACTSRODEPHNQALAKSKGLQYNNMSPLONNLHNOKTPPYDVSAMVPI  
AVMNGHDILADPOVAALPLRPLPLTHKEILPYNHLDFIWMADAPQEVNIVTM  
AED"

## CDS

BASE COUNT 396 a 289 c 282 g 379 t  
ORIGIN

Query Match 36.1%; Score 458.6; DB 12; Length 1346;  
Best Local Similarity 63.6%; Pred. No. 7.3e-123;  
Matches 698; Conservative 0; Mismatches 399; Indels 0; Gaps 0;

125 accgaagaagcattcatgaattatgaatcatccacaatacaagcgtacgtctgag 184  
132 ACCCTGAGCAACATGATGTTAGTCAGATGATACTTACGCGGATATCCAAAGTAGC 191  
185 aatataagtcgaactgaagaatgggtatatacctcttctgtaacaggtatcccgagcc 244  
192 AATATGAAGTGTACTGAAGATGCGTACATCTGGGGCTGTATGAATTCCTTATGGA 251  
245 tagtgaactaagaagaaggttccagcctgtgtgttactgaagcagcagtcagtgtg 304  
252 AGAAAAATTTCTAGATATACGGCAAGACCTGTGGCATATTTGCAAGCATGGTTGATTG 311  
305 gaggtgctagcagctggaattcccaacctgccacaataagcctggtcattccttcggag 364  
312 CATGACGCCAAACTGATTAACAAATCTGCCAAACAGCCTGGCCTTCATTCTACGCG 371  
365 atgctggttttgacgttgatggggaacagcagggaacacgcctggtctgcaaaacaca 424  
372 ATGCTGGCTATGATGTGTGGCGGAACAGTCGAGGAAATACATGTCGCCGAAAAAATG 431  
425 agaacctcccatagacaagaatgagttggtggtccttcagttatgatagatgagtcagtg 484  
432 TTTACTATTACACAGACTCAGTTGATCTGGCTTTCAAGCTTTGATGAATGAGCTAAT 491  
485 ttgacctctctgcagtgaataacttatttgcagaaaaacggcgcaaggaagacatcat 544  
492 ATGACCTTCCAGCCACATAGACTCATTTGTACAGAAAACTGCAAGGAAGATACACT 551  
545 atgctgctattacagagggcaccacatgggtcttattgcatcttccacatgcaagagc 604  
552 ATGTTGGTCACTCTCAGGCGACCACTATCGGTTTATGCTTTCTACCAATCCTGCTC 611  
605 tggctcagaataatcaaatgatttctgcttgcacccatagccactgttaagcagaa 664

Db 612 TGGCTAAAAAATCAAGAGTTTATGCAATTAGCTCCAGTTGCTACTGTGAAGTATACAG 671  
Oy 665 aaagcccgaggacaaatttltgtcgtccagatataatgatcaagagatgttgcga 724  
Db 672 AAAGTCCCTTTAAAAAGATTCTATTATTCCTAAAGTTCTTCTCAAGGTGATATTGGTA 731  
Oy 725 aaaagaatttctgtatcagaacagatttccagaacattgttatcccttctggtgc 784  
Db 732 ACAAAATGTTTCCATGCCCCAACATGATCAATCTTCTGGTGCGGAAGTGTCTCAG 791  
Oy 785 aggtatcttctgacagatttctgtagtaataatcatgatttcttctggtgattcaaccca 844  
Db 792 GGAAGCTGCTAGATTTCTCTGACCAACGCTTTATTTATCTTCTGTGATTTGCAAGA 851  
Oy 845 acaatataacatagcagcgaagtgatataatgctgcccacacacttctgtgaacatctg 904  
Db 852 AAAACTTAATGTGATGTGCTTGTATGTGATCTAGGCGCATATATCCAGAGAACATCTGA 911  
Oy 905 tgcagaatacttctacacttgagcagcaggaatgaaattctgtgtaactccgggcatctgac 964  
Db 912 CTCAGACCTTTTCCATGCGGACAGCTTGTAAATCTGGGAAGCTTCAAGCCTTAAGT 971  
Oy 965 ggggagagagagacccaataatctggaataatgcaatcaccactctctgtaaggtacag 1024  
Db 972 GGGGAAGTCCATTTACAGAACATGTTACACTACATCAACAAACGCGCTCCTACTATGATG 1031  
Oy 1025 tcaagatataacgcttccctcagaagaatgagcagcagcagcagcagcagcagcagc 1084  
Db 1032 TGTCCAGCATGACCTGCGCAATTCGATGTGTGAGAACGCTGCGCATCTCTGCTGATC 1091  
Oy 1085 cagaagacgtgaataatgctgctctctctgagagagcacaacctatctacataagaatctc 1144  
Db 1092 CCCAAGATGTGCGCATGCTGCTTCCCAAACTCCCAACCTTCTGTACCAATGAAGATTC 1151  
Oy 1145 ctgaatgggtcagcgtgatttcaatctggtgttggatgcttccatccgtatgtcaatg 1204  
Db 1152 TTCCTTCAATCCTCTGAGCTTCACTCATCTGAGGATGAGTGGCTTCAAGAGTTTACAAATG 1211  
Oy 1205 aaatcaccatctgagt 1221  
Db 1212 AGATGCTTACCATGATG 1228

RESULT 14  
AK009523  
LOCUS AK009523 1346 bp mRNA HTC 05-JUL-2001  
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310028109, full insert sequence.  
ACCESSION AK009523  
VERSION AK009523.1 GI:12844369  
KEYWORDS CAP trapper.  
SOURCE Mus musculus (strain: C57BL/6J) adult male tongue cDNA to mRNA, clone:2310028109.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1346)  
AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning methods in enzymology. 303, 19-44 (1999)  
JOURNAL MEDLINE 99279253  
PUBMED 10349636  
REFERENCE 2 (bases 1 to 1346)  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL MEDLINE 20499374  
PUBMED 11042159  
REFERENCE 3 (bases 1 to 1346)  
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,

**TITLE**  
RIKEN Integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multichannel sequencer  
Genome research. 10 (11), 1757-1771 (2000)

TITLE	REFERENCE	AUTHORS
4 (bases 1 to 1346)		
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.		
Functional annotation of a full-length mouse cDNA collection		
Nature 409, 685-690 (2001)		
5 (bases 1 to 1346)		
Adachi, J., Alizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,		

	Df	1032	TGTCAGCCATCGACCTGGCAATTGCAGTGTGGAA CGGTGCCCATC ATGACTTGCTGGTAGTT	1091
	Oy	1085	cagaagacgtgaataatgctgcctccttcaggcgfgaaccaacctcatccacaagaatatc	1144
	Dd	1092	CCCCAAGAATGTCGAATGCTGTCTCCCACAAC TCCACCACCTCTGTGCATAAGGAGTA TC	1151
	Oy	1145	ctgaattggaccacagttagttacttccccttgtygggtttgatgcctccacgtatgatcatacy	1204
	Dd	1152	TTCCCTAACAAATCACCTGGACCTTCATCCTGGGGATGTGA TGAGCGCCTCAAAGAGTTTTCAAT G	1211
	Oy	1205	aaatcatccaatcctgatgy 1221 	
	Dd	1212	AGATAGTTACCATGATG 1228 	
RESULT 15				
LOCUS AK009729		1346 bp mRNA HTc	05-JUL-2001	
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:J31004D0L3, full insert sequence.				
ACCESSION AK009729				
VERSION AK009729.1 GI:12844702				
KEYWORDS CAP trapper. Mus msculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone-lib;RIKEN full-length enriched mouse cDNA library clone:J31004D0L3.				
SOURCE Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
ORGANISM Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS Carninci,P. and Hayashizaki,Y.				
TITLE High-efficiency full-length CDNA cloning Methods in enzymology. 303, 19-44 (1999)				
JOURNAL MEDLINE 99279253				
PUBMED 10349636				
REFERENCE 2 (bases 1 to 1346) Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome research. 10 (10), 1617-1630 (2000)				
AUTHORS JOURNAL MEDLINE 20499374				
PUBMED 11042159				
REFERENCE 3 (bases 1 to 1346) Shibata,K., Itoh,M., Alzawa,K., Nagoaka,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,J., Nishi,K., Kitsumai,T., Tashiro,H., Itoh.M., Sumi.N., Ishii.Y., Nakamura.S., Hazama,M., Nishine.T., Harada,A., Yamamoto.R., Matsunoto.H., Sakaguchi.S., Ikegami.T., Kasbiwagi.K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuzura,S., Kawal.J., Kazakhi,I., Muramatsu,M., Inoue,Y., Kita.A. and Hayashizaki,Y. RIken integrated sequence analysis (RIS) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome research. 10 (11), 1757-1771 (2000)				
TITLE 4 (bases 1 to 1346) The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.				
JOURNAL MEDLINE Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)				
PUBMED 5 (bases 1 to 1346)				
REFERENCE Adachi,J., Alizawa,K., Akahira,S., Akimura,T., Anno,H., Arai,A., Hasekawa,T., Carninci,P., Fukuda,S., Furumushiji,Y., Futuno,M., Hanegaki,T., Harada,A., Hayatsu,N., Hiramoto,K., Hirooka,T., Horii,F., Imotani,R., Ishii,Y., ItoH,M., Izawa,M., Kato,H., Kawasaki,K., Kojima,Y., Kono,H., Koude,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Onho,M., Okazaki,Y., Orido,T., Owa.C., Saito,H., Saito,R., Sakai.C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takashiishi.F., Tanaka,Y., Tejima,Y., Toyota.T., Yamamura,T., Yasunishi,A., Yanuka,Y.,				

**TITLE**  
Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

**JOURNAL**  
Submitted (10-JUL-2000) Yoshihide Hayashizaki. The Institute of Direct Submission  
Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

**COMMENT**  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGACAGCAGACATCCCAAGACTCTTTTTCCTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGACAGCATCTCGAGTTAATTAAATACCCTCCCCCCC 3']. cDNA was cleaved with XhoI and SctI. Cloning sites, 5' end: XhoI; 3' end: SctI.  
Host: SOLR.

**FEATURES**  
**source**  
Location/Qualifiers.  
1..1346  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/db\_xref="MGD:MGI:1898390"  
/db\_xref="MGD:MGI:1914967"  
/clone="23100401L03"  
/sex="male"  
/tissue\_type="tongue"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
53..1240  
/note="putative"  
/cdon\_start=1  
/protein\_id="BAB26466.1"  
/db\_xref="GI:12844703"  
/translation="MMLIVTSVLSARCGAHGLFGKIGPKNPANMNVSOMITGYCP  
SEEEVEVDGGILGVTRIPRYGKNSSNICKRPYALQHELINSATWNTLPNSSLA  
FIADAGYEDVWLGNRSKNTNRKVNYSPDSVERMAFSEPMAXYDLPAIDTIVQT  
GOEKIHVVNGMSOGTTIGFIAFTSTPALKRKIFYALAPAYAKVYSPEPKXLSLPK  
FLKLVIENGMMFMFNHLYDOFLGTENVCSRLLDCSNALFICGFEPKMLNYSREPV  
YLIGNPGASTODLEFHAOLAKSKLOAYNMGSPLQMLHLNOKTPYOVUSAMTVDI  
AVNWNGHDILLADPDQDVAMLPRKPLENLILHKELIPYNILDPIMAMDAPQEYNNIYVTM  
AEI"

**CDS**

BASE COUNT	395 a	288 c	283 g	380 t
ORIGIN				
Query Match	36.1%	Score 458.6;	DB 12;	Length 1346;
Best Local Similarity	63.6%;	Pred. No. 7.3e-123;		
Matches 698;	Conservative 0;	Mismatches 339;	Indels 0;	Gaps 0;

```

Oy      125 accagaagcatcatgatataatagtgaacatcacacaacatcaaggctaccctgtgagg 184
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      132 ACCCTGAAGCAAACTGTAATGTGTAGTCAGATGAATACTTACTTGCGGATATCCAAGTAGG 191

Oy      185 aaattgaagtgcgaactgaaagatggatatcccttccttgtaaagaagattccctgagggc 244
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      192 AATATGAAGTGTACTCAAGATGGCTACATCTGGGGGTCTATATGAATTCCTTATGGGA 251

Oy      245 taftgcacaacctaaagaacaggtccagggcctgtgtygttaactgcagcatgacctagttg 304
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      252 AGAAAAATTCTGCNAATATATCGCAAGAGACCTGTGGCATTAATTGCAGATGCTTTGATFG 311

Oy      305 gaggtgtcgcaactgattccaacctgtcccacaacatagctgtggcttaattcttgacag 364
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      312 CATGACGCCCAAACTGGTTTCAATATGTGGCCAAACAACAGCTGCTGCCTTCAATTTCTACAG 371
    
```

OY	365	atgctgttttggacgtgtgatgtgggaaacagaagggaacgcctggtctcgtgaaacaca	424
Db	372	ATGCTGGCTATGATGTGTGGCTGGGAAACAGTCGAGGGATATCATGTGTCCGAAATAATG	431
OY	425	agaacacctcatagaccgaagatgaatctcgtggcttcagtatgatgtgtaagt	484
Db	432	TATACTTTTACACAGACTAGTGTGATTTCTGGGCTTTACGCTTTATATAATGGCTAAT	491
OY	485	ttgacctcttcgtcagtgatgaactatttttcgagaanaacggtccaggaaaagatctat	544
Db	492	ATGACCTTTCAGCGACACATATGACTTATGTATGAGAAACTGGACAAGAAATACACT	551
OY	545	atgtcgcgtattcaagggagaccacaatgtggtcttatgtcatttccacctgcgaagc	604
Db	552	ATGTTGGTCACTCTCAGGGGACACCATATGGGTTTATATTCCTTTCTACATCTGTCTC	611
OY	605	tggctcgaaatcaaatgatatttctgtcttgaccaccaatgacatgttaagctga	664
Db	612	TGGCTAAAAAATCAAGAGTTTATGCAATTAGCTTCAAGTCTACTGTGAGATACAG	671
OY	665	aaagccccggaccaaattttgtctgcgcagatatgaatgaatcaaggaatgtgtgca	724
Db	672	AAATCTCCTTAAAAAGATTTCATCTATTCCTATTCCTATTCCTTTCACAGGATATTGGTA	731
OY	725	aaaaagaattctgtatcagaccagaattctcagacaactgttatttaacctgtgtgc	784
Db	732	ACAAAATGTTCAATGCCCCACCAACTACTTATGATCAATTTCTTGGTACGGAAGTGTCTCAC	791
OY	785	aggtgatctcttgatcagattgttagaataatcatgttaactcttgggttgatccaacca	844
Db	792	GGGAGCGTCAGATCTTCTGTGACGACAGCTTATTCATCTCTGTGGATTGCAAGA	851
OY	845	acaatatgaacaatgagccgagcaagtgtatgtctgccacactctgttgaaactgt	904
Db	852	AAAACTTAAATGTGATGCTGTTGATGTGTATCTAGGGCATATTCACGACGAGAACATCTA	911
OY	905	tgcaaaatattctaacctgtgagccagcgagtgaaatctgtgtgaactccggtgattgact	964
Db	912	CTCAAGACCTTTTCCACTGTGGGACACAGCTTGCTPAAATCTGGAGAGCTTCAAGGCTTAACT	971
OY	965	gggggaattgagaccaaaatcttgaaaaatgtcaatcaagccaactcctgtaaagtacag	1024
Db	972	GGGGAATCTCATTAACAAMATGTTTCACTACATATGAGAAAGCGCTCCATCTATGATG	1031
OY	1025	tcgagagatgaagctgtccctcacagacaatgtgagcaggaagttaagactgtcttcaaat	1084
Db	1032	TGTGAGCCATGACCGTGCCAAATGGATGATGTGAACGCTGGCCATGATCATCTCGGTGATC	1091
OY	1085	cagaagacgtgaaaatgtctctctcgtgaggtgagccaacctcatctacaataaagtattc	1144
Db	1092	CCCAAGATGTGCAATGCTGCTTCCCAAATCTCCCAACTCTTGATACCATTAAGAGATTC	1151
OY	1145	ctgaatgggtcacggtgatctcatctgtgggttttgaatgtctcctcaacgtatgtacaatg	1204
Db	1152	TTCCCTCAATACACTGAGACTTATCTGTGGCGATGATCGCCTCAAGAGTTTTCATG	1211
OY	1205	aaatcatcatctgaatg 1221	
Db	1212	AGATAGTTTACCATGATG 1228	